languages

Exploring Cross-linguistic Effects and Phonetic Interactions in the Context of Bilingualism

Edited by

Mark Amengual

Printed Edition of the Special Issue Published in *Languages*

www.mdpi.com/journal/languages

MDPI

Exploring Cross-linguistic Effects and Phonetic Interactions in the Context of Bilingualism

Exploring Cross-linguistic Effects and Phonetic Interactions in the Context of Bilingualism

Editor

Mark Amengual

MDPI • Basel • Beijing • Wuhan • Barcelona • Belgrade • Manchester • Tokyo • Cluj • Tianjin

Editor
Mark Amengual
University of California
USA

Editorial Office
MDPI
St. Alban-Anlage 66
4052 Basel, Switzerland

This is a reprint of articles from the Special Issue published online in the open access journal *Languages* (ISSN 2226-471X) (available at: https://www.mdpi.com/journal/languages/special_issues/Cross-linguistic_Phoneti).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

LastName, A.A.; LastName, B.B.; LastName, C.C. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

ISBN 978-3-0365-0966-2 (Hbk)
ISBN 978-3-0365-0967-9 (PDF)

Cover image courtesy of Mark Amengual.

Contents

About the Editor

Mark Amengual is an Associate Professor of Applied Linguistics and the director of the UCSC Bilingualism Research Laboratory in the Department of Languages and Applied Linguistics at the University of California, Santa Cruz. His research interests focus primarily on experimental phonetics, bilingualism, and psycholinguistics. He has been the principal investigator or collaborator in several research projects on Spanish–Catalan bilinguals, Spanish–Galician bilinguals, Spanish–Portuguese bilinguals, Spanish heritage speakers in the United States, English heritage speakers and British expats in Spain, and Spanish–Otomi (Hñäñho) bilingual speakers in Mexico. This work has been published in journals, such as *Journal of Phonetics*, *The Journal of the Acoustical Society of America*, *Phonetica*, *Bilingualism: Language and Cognition*, *International Journal of Bilingualism*, *Linguistic Approaches to Bilingualism*, and *Applied Psycholinguistics*, among others.

languages

Editorial

Exploring Cross-Linguistic Effects and Phonetic Interactions in the Context of Bilingualism: Introducing the Special Issue

Mark Amengual

UCSC Bilingualism Research Laboratory, Department of Languages and Applied Linguistics, University of California, Santa Cruz, CA 95064, USA; amengual@ucsc.edu

1. Background and Motivation

Bilinguals who have acquired both of their languages simultaneously since birth or have learned their first language (L1) and their second language (L2) sequentially, as children or as adults, are able to produce and perceive two different sound systems. To accomplish this, a bilingual individual will have to attune their perceptual abilities to the sound system of each language and also produce acoustically distinct sounds depending on which language they are using. Over the past three decades, one of the central questions in bilingualism research has been to determine to which extent the bilingual's two phonological inventories are interconnected, and thus influence each other, and in what ways they remain independent, especially in comparison with the single system of monolinguals.

As predicted in current theoretical models of L2 phonological acquisition, such as the Speech Learning Model, SLM (Flege 1995); the revised Speech Learning Model, SLM-r (Flege and Bohn 2021); the Perceptual Assimilation Model, PAM (Best 1995); the Perceptual Assimilation Model of Second Language Learning, PAM-L2 (Best and Tyler 2007); and the Second Language Perception Model, L2LP (Escudero 2005), the learnability of L2 speech sounds is perceptual in nature and depends on the perceived phonetic distance between the sounds in the L2 and the most similar sounds in the L1 phonetic inventory. For instance, the SLM postulates that a bilingual's L1 and L2 sounds are interrelated and coexist in a common acoustic-phonetic space, with the bilingual sound system being a combination of the two languages' segmental inventories. If this is the case, the bilingual sound system will, by consequence, be prone to cross-linguistic influence.

Cross-linguistic phonetic/phonological influence has been defined as "the way in which a person's knowledge of the sound system of one language can affect that person's perception and production of speech sounds in another language" (Jarvis and Pavlenko 2008, p. 62). This type of cross-linguistic influence plays an important role in provoking the percept of a discernable 'foreign accent'. However, this accent is not exclusive to speech production. In addition to speaking with a "foreign" or L1-influenced accent, previous studies have also shown that language users also "hear with an accent" (Jenkins et al. 1995). A growing body of research on bilingual speech has examined cross-linguistic influence in the production, perception and processing of sounds in the L1 and the L2, focusing on how bilinguals categorize speech sounds, their sensitivity to phonetic variation, and the phonemic and phonetic abilities of these bilinguals in their production of language-specific segmental and suprasegmental features.

A number of these studies suggest that bilingual speakers do not necessarily produce acoustic targets free of influence from the L1 in their L2; rather, the bilinguals' combined or interrelated systems influence each other at a fine-grained acoustic level in speech production and perception. In order to better understand the phonological/phonetic systems of early and late bilinguals, it is clear that we are obligated to consider many variables in addition to age of acquisition, such as language proficiency, language dominance, language

Citation: Amengual, Mark. 2021. Exploring Cross-Linguistic Effects and Phonetic Interactions in the Context of Bilingualism: Introducing the Special Issue. *Languages* 6: 54. https://doi.org/10.3390/languages6010054

Received: 24 February 2021
Accepted: 5 March 2021
Published: 19 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

use, and other non-linguistic variables that are particular to the experiences of the bilingual individuals under investigation.

The present Special Issue brings together fifteen state-of-the-art papers that investigate phonetic interactions between the sound systems of multilingual speakers at the segmental and suprasegmental levels, using a wide variety of methodologies, and that inform current theoretical frameworks on bilingual speech. The authors are leading experts representing research institutions in the U.S., Canada, Australia, Spain, Poland, Korea, Germany, Austria, and the United Kingdom. The target languages that are investigated in this compilation of original research studies include English, Spanish, Japanese, Korean, Russian, German, French, and Catalan in the production and perception of early simultaneous bilinguals, heritage speakers, adult L2 learners, and L3 learners.

2. Article Summaries

The fifteen articles in this Special Issue are divided into two sections: five studies on speech perception and ten studies on speech production. Ideally, the reader should read the papers in the order established below, since this order renders a thematically coherent overview of the work included in the Special Issue. However, every paper may also be read independently and, as the reader will find out, the order of the papers on the journal website based on date of publication differs from the one adopted here.

Mariela López Velarde and Miquel Simonet (University of Arizona) open the bilingual speech perception section of this special issue, with the article entitled “The perception of postalveolar English obstruents by Spanish speakers learning English as a Foreign Language in Mexico”. In their study, López Velarde and Simonet investigate the perception of the English contrast /tʃ/–/ʃ/ (as in *cheat* and *sheet*) by two groups of L2 English learners who speak a different variety of Mexican Spanish. More specifically, this study compares the perceptual behavior of English learners from northwestern Mexico, where [tʃ] and [ʃ] are variants of the same phoneme in their Spanish variety with learners from central Mexico, whose native dialect includes only [tʃ]. The results of a word-categorization task and a categorical discrimination task show that both groups of learners have difficulties to identify and discriminate this contrast but the group of learners in northwestern Mexico find these tasks particularly challenging. These results show that phonetic variants in one’s native dialect modulate the perception of sounds of their second language.

The study by Charles B. Chang (Boston University) and Sungmi Kwon (Pukyong National University), “The contributions of crosslinguistic influence and individual differences to nonnative speech perception” investigates the relative importance of cross-linguistic transfer from a listener’s native language (L1) vis-à-vis individual learner differences. It explores the hypothesis that the nature of L1 transfer changes as learners gain experience with the L2, such that individual differences are more influential at earlier stages of learning and L1 transfer is more influential at later stages of learning. In a pretest-posttest design, novice L2 learners of Korean from diverse L1 backgrounds were examined with respect to their perceptual acquisition of novel L2 consonant contrasts: the three-way Korean laryngeal contrast among lenis, fortis, and aspirated plosives, and the /o/-/ʌ/ and /u/-/ɨ/ Korean vowel contrasts. The results indicate that even though pretest performance showed little evidence of L1 effects, posttest performance showed significant L1 transfer. These findings suggest that L1 knowledge influences L2 perception dynamically, as a function of the amount of experience that learners gain with the L2. Therefore, both individual differences and L1 knowledge play a role in L2 perception, but to different degrees over the course of L2 development.

The third study, “Perceived phonological overlap in second-language categories: The acquisition of English /r/ and /l/ by Japanese native listeners” by Michael D. Tyler (Western Sydney University), tests the idea that Japanese learners who have acquired English in a formal learning setting may perceive phonological overlap when they encounter L2 phones, and that the overlap may decrease with immersion experience in an L2 environment. In this study, Japanese native speakers differing in English L2 immersion and native English speakers completed a *forced category goodness rating task*, where they rated the goodness of fit of an auditory stimulus to an English phonological category label. The auditory stimuli consisted of 10 steps of a synthetic /r/-/l/ continuum, plus /w/ and /j/. The results indicated that less experienced Japanese participants rated steps at the /l/-end of the continuum as equally good versions of /l/ and /r/, but steps at the /r/-end were rated as better versions of /r/ than /l/, and for those with more than two years of immersion there was a separation of goodness ratings at both ends of the continuum. However, the separation was smaller than it was for the native speakers. These findings show that for L2 listeners the perceived phonological overlap appears to improve with immersion experience.

The fourth study on speech perception by Jaydene Elvin (California State University, Fresno), Daniel Williams (University of Potsdam), Jason Shaw (Yale University), Catherine Best (Western Sydney University), and Paola Escudero (Western Sydney University), "The role of acoustic similarity and non-native categorisation in predicting non-native discrimination: Brazilian Portuguese vowels by English vs. Spanish listeners" examines whether Australian English (AusE) and European Spanish (ES) listeners differ in their categorization and discrimination of Brazilian Portuguese (BP) vowels. The results from the perception tasks showed comparable performance for the AusE and ES participants in their perception of the BP vowels. More specifically, discrimination patterns were largely dependent on contrast-specific learning scenarios which were similar across AusE and ES, and the acoustic similarity between the individuals' own native productions and the BP stimuli were largely consistent with the participants' patterns of non-native categorization. Furthermore, the results indicated that both acoustic and perceptual overlap successfully predict discrimination performance. However, accuracy in discrimination was better explained by perceptual similarity for ES listeners and by acoustic similarity for AusE listeners. Finally, the group averages were found to explain discrimination accuracy better for ES listeners than predictions based on individual production data, but this was not the case for the AusE group. These findings are interpreted in light of the predictions put forth by the Second Language Perception Model, L2LP (Escudero 2005).

The final article in the speech perception section is also the only study in this special issue that investigates cross-linguistic effects in L3 acquisition, entitled "Cross-linguistic interactions in third language acquisition: evidence from multi-feature analysis of speech perception" by Magdalena Wrembel (Adam Mickiewicz University, Poznań), Ulrike Gut (University of Münster), Romana Kopečková (University of Münster), and Anna Balas (Adam Mickiewicz University, Poznań). This study contributes to the few studies that have examined cross-linguistic effects in the speech perception of multilingual learners by exploring the development of speech perception in a group of young L1 Polish speakers learning L2 English and L3 German. These multilingual individuals performed a forced-choice goodness task in their L2 and L3 to test their perception of rhotics and final obstruent (de)voicing. Data on their response accuracy and reaction times indicate that cross-linguistic influence in perceptual development is feature-dependent with relative stability evidenced for L2 rhotics, reverse trends for L3 rhotics, and no significant development for L2/L3 (de)voicing. These results also show that the source of cross-linguistic influence differed across the speakers' languages.

Shifting to bilingual speech production, the first study that opens this section is the article by Daniel J. Olson (Purdue University), "Short-term sources of cross-linguistic phonetic influence: Examining the role of linguistic environment". This production study investigates the potential for linguistic environment to serve as a source of short-term cross-linguistic phonetic influence. To test this, a group of L1 English-L2 Spanish bilinguals produced Spanish utterances in two different sessions in different locations: in an English-dominant linguistic environment in Indiana, USA, and in a Spanish-dominant linguistic environment in Madrid, Spain. The results from an acoustic analysis of VOT and native speaker global accent ratings show that the linguistic environment did not significantly impact either measure of phonetic production, regardless of a speaker's L2 proficiency.

These findings point to a possible primacy of the immediate context of an interaction, rather than broader community norms, in determining language mode and cross-linguistic influence.

The next study by Robert Mayr (Cardiff Metropolitan University), David Sánchez (Cardiff Metropolitan University), and Ineke Mennen (University of Graz), entitled "Does teaching your native language abroad increase L1 attrition of speech? The case of Spaniards in the United Kingdom" focuses on cross-linguistic influence from the L2 to the L1 (i.e., reverse transfer). Specifically, it examines the perceived L1 pronunciation of two groups of native Spaniards residing in the United Kingdom (Spanish language teachers and non-teachers) and monolingual Spanish controls in Spain. Crucially, the Spanish language teachers significantly used more Spanish at work than non-teachers. Global accentedness ratings were obtained from monolingual native Spanish listeners living in Spain, who rated short speech samples extracted from a picture-based narrative. The results showed significantly greater foreign-accent ratings for teachers than for non-teachers and monolinguals, and this non-native speech was associated with a range of segmental and suprasegmental features. These results suggest that language teachers who teach their L1 in an L2-speaking environment may be particularly prone to L1 attrition since they need to co-activate both of their languages in professional settings and are regularly exposed to non-native speech from L2 learners.

As in the previous study, the next article on bilingual speech production entitled, "The effect of instructed second language learning on the acoustic properties of First Language Speech" by Olga Dmitrieva (Purdue University), Allard Jongman (University of Kansas), and Joan. A. Sereno (University of Kansas) also contributes to the growing evidence of reverse transfer from the L2 to the L1 during L2 phonetic acquisition. In this study, Russian and English productions of 20 American classroom learners of Russian were compared to 18 English monolingual controls with a focus on the acoustics of word-initial and word-final voicing. The results indicated that learner's Russian voiced and voiceless stops were acoustically different from their English ones demonstrating a successful acquisition of these L2 Russian segments. These learners also showed an L1 phonetic change in comparison to the monolingual English speakers: their English VOTs were shortened, therefore, they were more Russian-like. This was taken as evidence of assimilation with Russian whereas the frequency of prevoicing in English was decreased, indicating dissimilation with Russian. With respect to word-final voicing, the duration of preceding vowels, stop closures, frication, and voicing during consonantal constriction all demonstrated drift towards Russian norms of word-final voicing neutralization. These findings demonstrate that L2-driven phonetic changes in the L1 are possible even in L1-immersed classroom language learners, challenging the role of reduced L1 use and highlighting the plasticity of the L1 phonetic system.

The next article by Magdalena Romera (Universidad Pública de Navarra) and Gorka Elordieta (Universidad del País Vasco-Euskal Herriko Unibertsitatea), entitled "Information-seeking question intonation in Basque Spanish and its correlation with degree of contact and language attitudes", analyzes the prosodic characteristics of Spanish in contact with Basque in the Basque Country. More specifically, the study focuses on the prosody of information-seeking yes/no questions, which present different intonation contours in Spanish and Basque. In contrast to previous work in urban areas, this study examined the suprasegmental features of speakers in rural areas. The results showed that falling intonational contours at the end of information-seeking absolute interrogatives were more common than in urban areas, and no correlation was found with the degree of contact with Basque and with attitudes towards Basque. The authors interpret these results as evidence that in rural areas the presence of Basque in daily life is stronger than in urban settings, and that there is a consolidated variety of Spanish used by all speakers regardless of their language attitudes. These findings reveal the relevance of subjective social factors in the degree of convergence between two languages.

Justin Davidson (University of California, Berkeley) in "Asymmetry and directionality in Catalan-Spanish contact: intervocalic fricatives in Barcelona and Valencia" examines another situation of language contact in Spain (Spanish/Catalan), this time focusing on the variable voicing and devoicing of intervocalic alveolar fricatives in Spanish, Barcelonan Catalan, and Valencian Catalan. The results from data elicited using a phrase-list reading task and sociolinguistic interviews reveal a stronger influence of Catalan on Spanish in Barcelona and Spanish on Catalan in Valencia, showing that these asymmetries, corroborated by attitudinal differences afforded to Catalan and Spanish in Barcelona and Valencia, reinforce the role of social factors in language contact outcomes.

The sixth article in this bilingual production section is entitled "Shared or separate representations? The Spanish Palatal Nasal in Early Spanish/English bilinguals" by Sara Stefanich (Northwestern University) and Jennifer Cabrelli (University of Illinois at Chicago). This acoustic study investigates whether a group of twenty Spanish heritage speakers living in the Chicagoland area have established a representation for the Spanish palatal nasal /ɲ/ (e.g., /kaɲon/ *cañón* 'canyon') that is separate from the similar, yet acoustically distinct English /n+j/ sequence (e.g., /kænjn̩/ 'canyon'). Duration and formant contour data elicited in a delayed repetition task in each language show that these early bilinguals distinguish between the Spanish /ɲ/ and English /ɲ+j/ in production, indicative of the maintenance of separate representations for these similar sounds. These results provide evidence of a lack of interaction between systems for bilinguals in this scenario.

In "Redefining sociophonetic competence: mapping COG differences in Phrase-Final fricative Epithesis in L1 & L2 speakers of French" by Amanda Dalola (University of South Carolina) and Keiko Bridwell (University of Georgia), the objective is to evaluate different measures of center of gravity (COG) in phrase-final fricative epithesis (PFFE) produced by L1 and L2 speakers of Continental French. Forty participants completed a reading task with target stimuli that elicited /i,y,u/ in phrase-final position. Results of the COG measures revealed that L2 speakers showed higher COG values than L1 speakers in low PFFE-to-vowel rations at the 25%, 50%, and 75% marks. After categorizing COG measures into six profile types on the basis of their frequencies at each timepoint (flat-low, flat-high, rising, falling, rising-falling, and falling-rising), the results revealed that although L1 speakers produced predominantly flat-low profile types at a lower percent devoicing, L2 speakers preferred multiple strategies involving higher levels of articulatory energy (rising, falling, rise-fall). These findings show that additional phonetic dimensions are necessary in the construct of L2 sociophonetic competence.

The next article by Laura Colantoni (University of Toronto), Ruth Martínez (University of Toronto), Natalia Mazzaro (University of Texas at El Paso), Ana T. Pérez Leroux (University of Toronto), and Natalia Rinaldi (University of Toronto) is entitled "A phonetic account of Spanish-English bilinguals' divergence with agreement". This study investigates the acoustic realization of Spanish word-final unstressed vowels /a,e,o/ in read and semi-spontaneous speech produced by 11 monolingual Spanish speakers, 13 early Spanish-English bilinguals, and 13 late Spanish-English bilinguals. The results of an acoustic analysis showed that early bilinguals exhibited clear patterns of vowel centralization and higher rates of hiatuses than the other two groups. Specifically, /a/ and /o/ were realized as centralized vowels, particularly with [+Animate] nouns. It is concluded that such variable vowel realizations may be a factor in the vulnerability to attrition in gender marking in Spanish as a heritage language.

The following article entitled "(Divergent) Participation in the California Vowel Shift by Korean Americans in Southern California" by Ji Young Kim (University of California, Los Angeles) and Nicole Wong examines the participation in the California Vowel Shift by Korean Americans in Los Angeles. First generation, generation 1.5, and second generation Korean-Americans are compared to Anglo-Californians and non-immigrant Korean late learners of English with respect to their English vowel productions. Results from a picture narrative task show a clear distinction between early vs. late bilinguals: while the first-generation Korean Americans and the late learners showed apparent signs of Korean influence, the 1.5- and second-generation Korean Americans participated in most patterns of the California Vowel Shift. However, divergence from the Anglo-Californians was observed in early bilinguals' speech. These findings indicate that age of arrival has a strong effect on immigrant minority speakers' participation in local sound change and that second-generation Korean Americans may be in a more advanced stage of the California Vowel Shift than Anglo-Californians or that the California Vowel Shift is on a different trajectory for these speakers.

The final article of the bilingual speech production section, and of the Special Issue as a whole, is entitled "Interlingual interactions elicit performance mismatches not "compromise" categories in early bilinguals: Evidence from meta-analysis and coronal stops" by Joseph V. Casillas (Rutgers University). This study uses meta-analytic techniques and coronal stop data from early bilinguals in order to assess the claim that early bilinguals produce the sounds of their languages in a manner that is characterized as "compromise" with regard to monolingual speakers. In this paper, Casillas provides an assessment of the literature and presents an acoustic analysis of coronal stops from early Spanish-English bilinguals. A range of studies were coded for linguistic and methodological features, as well as effect sizes, and then analyzed using a cross-classified Bayesian meta-analysis, and the results indicated that the pooled effect for "compromise" VOT was negligible. The acoustic analysis of the coronal stops showed that a group of early Spanish-English bilinguals often produced Spanish and English targets with mismatched features from their other language, and that these are likely to have occurred as a result of interlingual interactions elicited by the experimental task. Taken together, these results are interpreted as evidence that early bilinguals do not have "compromise" VOT, though their speech involves dynamic phonetic interactions that can surface as performance mismatches during speech production.

3. Final Remarks

The articles contained in this Special Issue examine cross-linguistic phonetic/phonological influence in bilinguals and trilinguals. These experimental studies contribute to the field with novel empirical data collection techniques, sophisticated methodologies and acoustic analyses, and present findings with robust theoretical implications for a variety of subfields, such as L2 acquisition, L3 acquisition, Laboratory Phonology, Acoustic Phonetics, Psycholinguistics, Sociophonetics, Bilingualism, and Language Contact. These studies will serve as a source of motivation for future research and to further elucidate the nature of phonetic interactions in the context of bilingualism and multilingualism.

I would like to thank the authors for submitting their research to this Special Issue. It has been a privilege to have worked with each and every contributor to this volume. Last but not least, I would also like to express my gratitude to the following reviewers, whose expertise, thorough evaluation, constructive feedback, and attention to detail greatly helped to improve the articles included in this Special Issue: A. Raymond Elliot (University of Texas at Arlington), Amanda Boomershine (University of North Carolina Wilmington), Anabela Rato (University of Toronto), Anel Brandl (Florida State University), Antonio Romano (Università degli Studi di Torino), Avizia Long (San José State University), Becky Muradás-Taylor (York St. John University), Brandon Baird (Middlebury College), Brendan Regan (Texas Tech University), Chiara Celata (Università degli Studi di Urbino Carlo Bo), Christine Shea (University of Iowa), Daniel J. Olson (Purdue University), Denise Osborne (SUNY Albany), Eivind Nessa Torgersen (Norwegian University of Science and Technology), Elena Schoonmaker-Gates (Elon University), Elisabeth Mayer (Australian National University), Erik Thomas (North Carolina State University), Esther De Leeuw (Queen Mary University of London), Fernando Llanos (University of Texas at Austin), Francesc Roca (Universitat de Girona), Germán Zárate-Sández (Western Michigan University), Gillian Lord (University of Florida), Isabelle Darcy (Indiana University), Katharina Schuhmann (The Pennsylvania State University), Laura Spinu (Kinsborough Community College, CUNY), Lucrecia Rallo Fabra (Universitat de les Illes Balears), Markus Christiner

(Universität Wien), Melinda Fricke (University of Pittsburgh), Michael Gradoville (Arizona State University), Miquel Llompart (Friedrich-Alexander-Universität Erlangen-Nürnberg), Mónica Chamorro (Universidad Javeriana Cali), Robert Mayr (Cardiff Metropolitan University), Sara Zahler (SUNY Albany), Silvina Bongiovanni (Michigan State University), Timothy Face (University of Minnesota), and Zsuzsanna Fagyal (University of Illinois Urbana-Champaign).

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

References

Best, Catherine. 1995. A direct realist perspective on cross-language speech perception. In *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*. Edited by Winifred Strange. Timonium: York Press, pp. 171–204.

Best, Catherine T., and Michael Tyler. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In *Language Experience in Second Language Speech Learning, in Honor of James Emil Flege*. Edited by Ocke-Schwen Bohn and Murray J. Munro. Amsterdam: John Benjamins, pp. 13–24.

Escudero, Paola. 2005. *Linguistic Perception and Second Language Acquisition*. Utrecht: Utrecht University.

Flege, James E. 1995. Second-language speech learning: Theory, findings and problems. In *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*. Edited by Winifred Strange. Timonium: York Press, pp. 229–73.

Flege, James E., and Ocke-Schwen Bohn. 2021. The revised Speech Learning Model. In *Second Language Speech Learning, Theoretical and Empirical Progress*. Edited by Ratree Wayland. Cambridge: Cambridge University Press, pp. 3–83.

Jarvis, Scott, and Aneta Pavlenko. 2008. *Crosslinguistic Influence in Language and Cognition*. New York: Routledge.

Jenkins, James J., Winifred Strange, and Linda Polka. 1995. Not everyone can tell a "rock" from a "lock": Assessing individual differences in speech perception. In *Assessing Individual Differences in Human Behavior: New Concepts, Methods, and Findings*. Edited by David J. Lubinksi and René V. Dawis. Palo Alto: Davies-Black Publishing, pp. 297–325.

languages

Article

The Perception of Postalveolar English Obstruents by Spanish Speakers Learning English as a Foreign Language in Mexico

Mariela López Velarde and Miquel Simonet *

Program in Second Language Acquisition and Teaching, University of Arizona, Tucson, AZ 85721, USA; marialopez@arizona.edu

* Correspondence: simonet@arizona.edu

Received: 10 April 2020; Accepted: 15 June 2020; Published: 22 June 2020

Abstract: The present study deals with the perception (identification and discrimination) of an English phonemic contrast (/tʃ/–/ʃ/, as in *cheat* and *sheet*) by speakers of two Mexican varieties of Spanish who are learning English as a foreign language. Unlike English, Spanish does not contrast /tʃ/ and /ʃ/ phonemically. Most Spanish varieties have [tʃ], but not [ʃ]. In northwestern Mexico, [ʃ] and [tʃ] find themselves in a situation of "free" variation—perhaps conditioned, to some extent, by social factors, but not in complementary distribution. In this variety, [ʃ] and [tʃ] are variants of the same phoneme. The present study compares the perceptual behavior of English learners from northwestern Mexico, with that of learners from central Mexico, whose native dialect includes only [tʃ]. The results of a word-categorization task show that both groups of learners find *cheat* and *sheet* difficult to identify in the context of each other, but that, relative to the other learner group, the group of learners in northwestern Mexico find this task to be particularly challenging. The results of a categorical discrimination task show that both learner groups find the members of the /tʃ/–/ʃ/ contrast difficult to discriminate. On average, accuracy is lower for the group of learners in northwestern Mexico than it is for the central Mexicans. The findings suggest that the phonetic variants found in one's native dialect modulate the perception of nonnative sounds and, consequently, that people who speak different regional varieties of the same language may face different obstacles when learning the sounds of their second language.

Keywords: second language acquisition; phonology; discrimination; cross-linguistic assimilation; obstruent; affricate; fricative; dialect; English; Spanish

1. Introduction

Most people "have an accent" when speaking a language other than their native one(s). This has been widely documented, and we currently have a sizeable scientific literature describing and explaining this phenomenon—see the following reviews (Best and Tyler 2007; Bohn 2017; Broselow and Kang 2013; Chang 2019; Colantoni et al. 2015; Davidson 2017, p. 201; Eckman 2012; Flege 1995; Piske et al. 2001; Simonet 2016). Interestingly, "having an accent" is not restricted to speech production, but also manifested in perception. Current models of L2 speech acquisition account for those findings—typically from the perspective of perception and categorization—by postulating some sort of interaction between native and nonnative sounds in the representational network of bilinguals (Best and Tyler 2007; Escudero 2005; Flege 1995; van Leussen and Escudero 2015). L2 learners have an "accent" in their L2, these models state, because they already have internalized knowledge of a first language (L1). Native and nonnative sounds must find a way to co-exist, and this typically results in modifications to the nature of such sounds. In other words, L2 listeners assimilate the sounds of their L2 in terms of the categories that are robustly represented in their phonology by the time they are learning the L2 (i.e., their L1), and they acquire these new sounds as a function of how they map them.

In English, /tʃ/ and /ʃ/ constitute a phonemic contrast, as seen in minimal pairs such as *sheet–cheat* and *chair–share*. Spanish does not have this contrast. Most varieties of Spanish have [tʃ] in their inventory, but they do not have [ʃ] (Hualde 2005, pp. 152–72). In spelling, /tʃ/ is systematically represented by the digraph <ch>, as in *charco* 'puddle' [ˈtʃaɾko] and *chamarra* 'jacket' [tʃaˈmara], and most Spanish speakers would consistently pronounce this phoneme as a postalveolar affricate, [tʃ]. It follows that, if they are to acquire the English /ʃ/–/tʃ/ contrast successfully, native Spanish speakers who possess this particular phonological system must develop a new phoneme, (/ʃ/), in opposition to one they can recycle from their native language, (/tʃ/); they must create a new contrastive category, and they must assign to it a new phonetic substance. Learning new sounds and new oppositions typically presents a significant phonological challenge (Best and Tyler 2007; Colantoni et al. 2015; Escudero 2005).

Native speakers of some regional varieties of Spanish, on the other hand, may have an acquisitional obstacle of a different nature. In some dialects, both [tʃ] and [ʃ] are found, but not in phonemic opposition. One such variety is spoken in northwestern Mexico, where people are known to pronounce Spanish words that have <ch> variably, with either [ʃ] or [tʃ] (Alessi Molina and Díaz 1994; Amastae 1996; Brown 1989; Carreón Serna 2007; Martín Butragueño 2009; Méndez 2017; Moreno de Alba 1994; Serrano Morales 2000, 2009). In northwestern Mexico, therefore, [ˈtʃaɾko] and [ˈʃaɾko] are common variants of the same word, *charco* 'puddle.' It seems to follow that, in order to acquire the English /ʃ/–/tʃ/ contrast successfully, native Spanish speakers from northwestern Mexico do not need to learn any new sounds. They already have both [tʃ] and [ʃ] in their inventory of phonetic categories. However, and this might be crucial, what they must do is learn that these two sounds are not variants of the same phoneme, like they are in their native Spanish variety, but separate phonemes (or separate contrastive categories). Learning new mappings between surface and underlying phonological representations presents a substantial acquisitional obstacle of a different kind (Barrios et al. 2016b).

The present study aims at contributing to the literature on the effects of native linguistic experience on the acquisition of L2 sounds. Most importantly, it examines the relative difficulty of developing new categories (new sounds) *versus* that of developing new phonemic contrasts between sounds one can reuse from one's native phonetic inventory (new mappings). The study is concerned with categorization patterns in the perception of an English phonemic contrast, /ʃ/–/tʃ/, by two groups of L1 Spanish learners of English who speak different regional varieties of their native language.

1.1. Cross-Linguistic Interactions in L2 Speech Acquisition

The fact that native and nonnative sounds interact in the bilingual mind is illustrated by a well-known example, that of Spanish-speaking learners of English, who tend to have difficulties with the English /i/–/ɪ/ and /æ/–/ɑ/ contrasts (Barrios et al. 2016a; Casillas 2015; Escudero and Boersma 2004; Flege et al. 1994, 1997; Flege and Bohn 1989; Kondaurova and Francis 2008; Morrison 2008, 2009). Spanish has five phonemic vowels, /i e a o u/, and Spanish-speaking learners of English tend to assimilate both English /i/ and /ɪ/ to a single native Spanish vowel, /i/ (e.g., Flege and Bohn 1989). This two-to-one cross-linguistic assimilation pattern creates an acquisitional obstacle for this learner population because it makes the two members of the English /i/–/ɪ/ contrast difficult to discriminate from each other (e.g., (Flege et al. 1994)). The English /ɑ/–/æ/ contrast also presents a challenge for Spanish-speaking learners, as both English vowels are cross linguistically assimilated to a single Spanish vowel, /a/ (Barrios et al. 2016a; Casillas and Simonet 2016). These findings indicate that the obstacles L2 learners encounter when acquiring the phonology of their L2 are, at least in part, determined by the listeners' native language background and the cross-linguistic assimilations established between L1 and L2 sounds.

Several theoretical accounts have attempted to explain the obstacles learners face during their acquisition of the L2 phonology. Two such models are the Perceptual Assimilation Model applied to L2 learning (PAM-L2) (Best and Tyler 2007) and the Second Language Linguistic Perception model

(L2LP) (Escudero 2005; van Leussen and Escudero 2015). Both of these frameworks postulate that the native and nonnative sounds of L2 learners interact (in some way). The manner in which this interaction is modelled, however, differs in the two accounts. The PAM-L2 proposes that difficulties arise as a function of the assimilability of L2 contrasts to L1 categories. Cross-linguistic assimilation is claimed to rely on the phonetic (in particular, articulatory) similarity between the L2 and the L1 sounds. The following are three of the possible assimilation patterns the PAM-L2 operationalizes: (i) When two L2 phones are cross linguistically assimilated or equated to two different L1 phonemes, a *two-category* assimilation (TC) is said to have occurred. In a TC scenario, discrimination of the two key L2 segments is predicted to be excellent, since the discrimination of the two corresponding L1 categories is assumed to be optimal. (ii) In contrast, if two L2 sounds are cross linguistically assimilated to the same L1 category and both are equally similar to the L1 sound, a *single-category* assimilation (SC) pattern occurs. The model predicts that, in cases of SC assimilation, the discrimination of the two key L2 phones is poor, as the two L2 sounds will have been categorized as variants of the same sound. This type of cross-linguistic assimilation pattern is particularly challenging for learners. (iii) A third type of assimilation pattern is called *category-goodness* assimilation (CG). In a CG pattern, two contrastive L2 sounds are assimilated to the same L1 category, but the cross-linguistic similarity is greater for one of the categories than for the other. In such a situation, discriminating between two key L2 categories is predicted to range from moderate to good, depending on the degree of category-goodness assimilation for each of the L2 segments.

The L2LP model differs from the PAM-L2 in some important ways. The L2LP claims that, at the initial stages of the L2 learning process, learners transfer or duplicate the entire L1 system to form an interlanguage system. Although the L2 system begins as a duplicate of the L1 grammar (a transferred grammar), this only occurs once, and it is subsequently handled as a separate phonological grammar. The novel system is equipped with the same learning mechanisms available in the L1, and it evolves as experience with the L2 increases. Technically, therefore, the L2LP rejects that the L1 and L2 phonological systems interact because it rejects that the two reside in a common representational network. Nevertheless, since the dedicated L2 system begins its course as an exact copy of the L1, the sound categories and mapping strategies learners developed for the L1 powerfully determine the manner in which L2 sounds are processed, perceived, produced, represented, and, ultimately, learned (or not learned). An aspect of the L2LP that resembles the PAM-L2 is that it operationalizes the existence of "cross-linguistic" *comparisons* in terms of L1 and L2 contrasts based on the phonetic (acoustic, in this case) similarity of L1 and L2 sounds. For instance, the cross-linguistic comparison described as single-category assimilation (SG) in the PAM-L2 is called *new scenario* in the L2LP, and both models predict that learning the L2 contrast in this particular scenario is challenging. What the PAM-L2 calls a two-category (TG) assimilation pattern, the L2LP calls a *similar scenario*; that is, when two L2 sounds each resemble a different L1 sound, learning these two categories is predicted to be easy. In sum, while some aspects of the PAM-L2 and the L2LP make them substantially different from each other, other principles of the two models are fundamentally identical.

Creating a new category during L2 acquisition is particularly difficult in cases in which two contrastive categories of the L2 are cross linguistically assimilated to a single category. A (perhaps) different kind of phonological obstacle presents itself when learners must develop new phonological mappings affecting sounds that already exist in their native inventory. For instance, Spanish has both [d] and [ð], but these are in complementary distribution—the two sounds are allophones (variants) of the same contrastive category, /d/. In English, on the other hand, these two sounds are contrastive, as illustrated by the minimal pair *den–then*. It follows that Spanish-speaking learners of English must develop new mappings between surface sounds that already exist in their inventory and new underlying representations (Barrios et al. 2016b). In other words, they must learn that two sounds that are linked to a single phonemic representation in their L1 are actually contrastive in their L2—they are linked to separate phonemes in the L2. It has been hypothesized that this type of acquisitional obstacle, called *allophonic split*, is particularly challenging for L2 learners (Eckman et al. 2001; Lado 1957).

This prediction derives from the finding that, in native speech, discriminating between contrastive categories is much easier than doing so between categories that are phonetically distinct but not contrastive; in other words, sounds that are not contrastive are perceived to be more similar to each other than contrastive sounds are (Barrios et al. 2016b; Johnson and Babel 2010). The literature on this learning scenario is scant but, interestingly, a recent study has shown that an obstacle of the kind described here does not "cause consistent difficulty for advanced L2 learners in perception" (Barrios et al. 2016b, p. 14). At this juncture, therefore, it is not known which of the learning scenarios—learning a new sound *versus* learning a new mapping—presents a greater challenge.

The present study is singularly placed to compare the relative difficulty of two of the learning scenarios discussed above: (i) the need to acquire a new sound category (that is, a new phoneme together with a new surface allophone), and (ii) the need to acquire a new mapping between surface and underlying representations (that is, a new phoneme for an already existing surface allophone).

1.2. Regional Dialects and L2 Speech Acquisition

The present study compares the perceptual behavior of two groups of Spanish-speaking learners of English. The two groups of learners differ in their region of origin; that is, they were brought up as speakers of two different geographical varieties of Mexican Spanish. The premise of our study is that the particular, specific L1 experience of L2 learners can determine, to some extent, the obstacles they encounter (and progression paths they take) when learning their L2. It follows that the phonology of the native dialect can modulate the acquisition of the phonology of the L2. A handful of recent studies have examined the potential role of regional dialect on L2 development. Some have explored the acquisition of different L2 dialects, that is, how people who speak the same language but are learning different varieties of the L2 differ in their linguistic behavior (Baker and Smith 2010; Escudero and Boersma 2004). Others have analyzed the potential effects of the native dialect on the acquisition of the L2, that is, how people who speak different varieties of the same language progress towards learning the same L2 (Chládková and Podlipský 2011; Escudero et al. 2012; Escudero and Chládková 2010; Mayr and Escudero 2010).

It has been demonstrated that people who speak the same native language but are exposed to different regional varieties of their L2 can face different cross-linguistic assimilation scenarios, leading to potentially different learning paths. For instance, Escudero and Boersma (2004) examined how two groups of Spanish-speaking learners of English perceived the English /i/–/ɪ/ contrast. One of the groups was learning English in Scotland, whereas the other was hypothesized to have been exposed mostly to the variety spoken in the South of England. This study found that learners in the two exposure groups behaved differently in their vowel categorization patterns. The authors attributed this to the acoustic properties of the particular target vowels involved, which led to different cross-linguistic assimilation patterns.

One's native dialect also seems to modulate the cross-linguistic assimilation patterns one will establish. For instance, Escudero et al. (2012) investigated the perceptual assimilation patterns displayed by Dutch-speaking L2 learners of English. It is well known that speakers of Dutch tend to have difficulties with the acquisition of the English /æ/–/ɛ/ contrast. Interestingly, Escudero and colleagues noted that, since the acoustics of the vowel categories of two regional varieties of Dutch (North Holland and Flanders) differ, it would be reasonable to predict diverging patterns of cross-linguistic assimilation for learners of English living in these two regions. Indeed, it was found that differences in the native-dialect phonetic system led to differences in cross-linguistic assimilation of vowels for learners in these regions. There is ample evidence that native phonology determines, to some extent, the learning paths of people acquiring a L2; evidently, the term "native phonology" refers to the individual phonological system of a given learner—their native phonological competence, which is based on their personal linguistic experience—and not to that of the "standard" dialect of a given learner's L1.

The present study compares the perceptual behavior of speakers of two regional varieties of Mexican Spanish when confronting an acquisitional challenge, the English /ʃ/–/tʃ/ contrast. Dialectological descriptions of the native phonology of speakers of these two regional varieties suggest differences that could lead to diverging patterns of cross-linguistic assimilation between their native consonants and those of English. This could lead to differences in their acquisitional obstacles and phonological learning paths.

1.3. Postalveolar Obstruents in Mexican Spanish

Dialectologists identify four regional varieties of Spanish in Mexico: central, coastal, northern, and peninsular (Yucatán) (Lope Blanch 1990; Martín Butragueño 2011, 2014; Moreno de Alba 1994). One of the phonological variables used to map the regional varieties of Mexican Spanish concerns the pronunciation of the first consonant in words such as *charco*, 'puddle', and *chamarra*, 'jacket.' In most dialects of Spanish, both in the Americas and in the Iberian Peninsula, this consonant is pronounced as a postalveolar affricate. This is true of most varieties of Mexican Spanish as well. Therefore, in most regions in Mexico, including the central highlands (the socially prestigious regional variety), *charco* 'puddle' is pronounced as [ˈtʃaɾko]. The speech of people born and raised in the northwestern Mexican states—including Sonora, Chihuahua, and Baja California Norte, among others—is characterized by a pattern of phonetic variation in which the postalveolar obstruent in *charco* 'puddle' may be pronounced as either [tʃ], an affricate, or [ʃ], a fricative. Therefore, in northwestern Mexican speech, *charco* 'puddle' is pronounced sometimes as [ˈtʃaɾko] and sometimes as [ˈʃaɾko]. Although variationist studies vary significantly in their reported fricativization rates ("fricativition" refers to the practice of pronouncing <ch> as [ʃ], a diachronic innovation), what seems clear is that, in this Spanish dialect, both the fricative and affricate variants of this variable are found (Alessi Molina and Díaz 1994; Brown 1989; Carreón Serna 2007; Méndez 2017).

Note that the use of the two variants of <ch> is not determined by a phonological rule; in other words, the two variants are not in complementary distribution, but in a scenario of "free" variation. In reality, the variation is not completely free: the investigations that have explored the phonetic variation that affects the pronunciation of <ch> have identified a number of social factors that may modulate to some extent variant choice. Among the social factors involved are age, level of education, and gender. Studies vary in their reported effects of gender (Carreón Serna 2007; Méndez 2017), and some claim that gender is meaningful only when it interacts with age (Jaramillo and Bills 1982). Age might also be relevant only when correlated with level of education (Jaramillo and Bills 1982). At any rate, what is important for our present purposes is that the alternation between [tʃ] as [ʃ] is not determined by a phonological process. The two sounds are neither contrastive nor in complementary distribution, since the same lexical item may be pronounced with either variant.

The constant exposure to the variation that affects <ch> has been found to affect northwestern, Mexican Spanish listeners' patterns of spoken word recognition. In a lexical access investigation, López Velarde and Simonet (2019) confirmed that listeners in northwestern Mexico are equally likely to accept Spanish word forms produced with either variant of <ch>. They also found that both variants are equally likely to prime listeners for the efficient recognition of spoken words (that is, [ˈʃaɾko] primes [ˈtʃaɾko] as much as [ˈtʃaɾko] does), which suggests that this group of listeners store Spanish words with both variants within the same abstract (or prototypical) mental representation. This study confirms that the two variants of <ch> are indeed allophones of the same phoneme. These findings suggest that people who experience sociophonetic variability in their speech community may store more than one phonetic variant in their long-term mental representation of words.

1.4. The Present Study

The current study focuses on a phonemic contrast of English—that between /tʃ/, as in *cheat*, and /ʃ/, as in *sheet*—and investigates the perceptual identification and discrimination patterns pertaining to this contrast displayed by two groups of L2 learners of English whose native language is Spanish. Our learner sample was recruited from two dialectal regions, central and northwestern Mexico. More specifically, this study explores how speakers of northwestern Mexican Spanish, who are recurrently exposed to the sociophonetic variability that affects Spanish <ch> in their speech community, perceive the target English contrast (/ʃ/–/tʃ/), and how these perceptual habits differ (if at all) from those demonstrated by speakers of central Mexican Spanish, who lack experience with this specific variability pattern.

Even though we hypothesize that both of our target populations of L2 learners of English may find this contrast relatively difficult to master, we believe that the specific learning obstacles the two populations experience are different—and this, we hypothesize, results in different learning outcomes. On the one hand, we postulate that the English /ʃ/–/tʃ/ contrast will prove to be relatively difficult for central Mexican learners because [ʃ], as in *sheet*, does not correspond to any sound in their dialect of Spanish while [tʃ], as in *cheat*, does. It is possible that these learners assimilate the two phonemes of English to the same native category, /tʃ/. Nevertheless, since [tʃ] and [ʃ] are phonetically quite distinct, and since central Mexican Spanish has both affricates (/tʃ/) and fricatives (/s, f, h/), it could be the case that (adopting PAM's terminology) the English /ʃ/–/tʃ/ contrast presents a category-goodness (CG) assimilation pattern, one in which English /tʃ/ is assimilated to central Mexican Spanish /tʃ/ with a very high goodness of fit and English /ʃ/ also assimilates to this central Mexican Spanish phoneme but with a lower goodness of fit.

Northwestern Mexican Spanish speakers, unlike people from central Mexico, are exposed to two variants of <ch> in their native dialect, [tʃ] and [ʃ]. These are two allophones of the same phoneme. For this reason, we hypothesize that speakers of northwestern Mexican Spanish are likely to assimilate both English /tʃ/ and English /ʃ/ to the same native phoneme, and that the goodness of fit of these two cross-linguistic assimilation patterns is likely to be similarly high. This might create (adopting PAM's terminology again) a single-category (SG) assimilation pattern. Since CG assimilation patterns are expected to lead to better discriminability than SC ones (Best et al. 1988, 2001; Best and Tyler 2007), we hypothesize that the error rates in the identification and discrimination of our target English contrast will be larger for learners in northwestern Mexico than for central Mexican learners.

An alternative way to frame the learning scenario for the northwestern Mexican Spanish speakers is that of an allophonic split (Barrios et al. 2016b; Eckman et al. 2001). These learners must unlearn that both [tʃ] and [ʃ] are mapped onto the same phoneme (as they are in their native dialect of Spanish), and they must develop a new phoneme specific to the L2 to which only one of these two phonetic categories is mapped. In other words, speakers of northwestern Mexican Spanish must develop a new phonological, underlying category and remap their sound categories so that they are each assigned to a different contrastive unit. Are the northwestern Mexican Spanish speakers more or less likely than the central Mexicans to succeed in their discrimination of the members of English /ʃ/–/tʃ/ contrast? This is the fundamental research question that motivates the present study.

To compare the acquisition of our target English phonemic contrast with a contrast about which much is known, we selected a second phonemic contrast of English to serve as a control condition, that between /i/ and /ɪ/. It is well known that native speakers of Spanish find the *seat–sit* contrast very difficult to discriminate and, therefore, learn (Casillas 2015; Escudero and Boersma 2004; Kondaurova and Francis 2008; Morrison 2008, 2009); they also find the two members of the contrast very difficult to identify against each other. Therefore, the *seat–sit* contrast, tested in our experiments alongside our target contrast, *sheet–cheat*, serves as a control condition, one that should be similarly challenging for both of our target learner populations.

2. Method

2.1. Participants

The data were collected in two locations in Mexico, Hermosillo and Santiago de Querétaro, which are the two largest cities in the states of Sonora and Querétaro, respectively. The participants in Hermosillo were lifelong residents of the state of Sonora. The majority of the participants had lived in Hermosillo from birth, and those born in other municipalities had moved to the city as children. Many of the participants tested in Santiago de Querétaro were not born in the city of Santiago, but they reported having moved there as children or as teenagers. The Querétaro residents in our sample who were not natives to Querétaro were born in other central states of the country, such as Guanajuato, Jalisco, Morelos, and Puebla. Particularly with respect to their treatment of /tʃ/, as well as to that of many other sounds, the central highlands of Mexico form a single dialectal area. In sum, data were collected in two dialectal areas, the northwest (exemplified by Hermosillo, Sonora) and the central highlands (exemplified by Santiago de Querétaro, Querétaro).

A total of 88 people (44 from Sonora, 44 from Querétaro) participated in this study. Participants' ages ranged between 18 and 43 years old. All but five participants were college students at the time of testing, graduate or undergraduate. Three participants had graduated with a college degree, and two had not completed college and were working in the industry. The high number of college students or college graduates in our sample is due to our having recruited our participants in college settings, the *Universidad de Sonora* (Hermosillo) and the *Universidad Autónoma de Querétaro* (Santiago de Querétaro). The educational and professional profile of the participants is not fully representative of the general population native to these locations—highly educated people are overrepresented. The profile, however, might be representative of the narrower population, in these locations, who have learned English as a foreign language. At any rate, the social profile of the two dialectal groups does not differ—both groups consist of highly educated people who are learning English as a foreign language in a school setting. All participants study (or studied) English in college.

Participants responded to the Bilingual Language Profile questionnaire (Gertken et al. 2014) The questionnaire collects information regarding the listeners' linguistic background and L2 learning experience with a focus on attitudes, history, self-assessed proficiency, and daily usage of the two languages. The questionnaire produces a language dominance score along a spectrum centered around 0, which represents balanced bilingualism. The participants in our study are expected to be Spanish dominant; in our implementation of the survey, dominance in Spanish is captured with scores ranging between 0 and −218.

We also administered an English vocabulary-size test to assess the participants' English knowledge, the LexTALE. The LexTALE (www.lextale.com) is a standardized test designed to measure vocabulary size in language learners (Lemhöfer and Broersma 2012). To the extent that vocabulary size reflects overall knowledge of the language, the LexTALE provides an indicator of a person's knowledge of English. It seems reasonable to speculate that acquisition of phonemic, contrastive categories is based upon vocabulary knowledge (Simonet 2016). The test consists of 60 trials, comprising 40 English words and 20 nonwords, and these are presented to participants for them to make lexical decisions on. The resulting score is expressed in percent-correct units, and it is corrected for the unequal number of words and nonwords. In this study, the test was administered using PsychoPy 2 (Peirce et al. 2019). After responding to the BLP and the LexTALE, participants proceeded to complete the identification task followed by the discrimination task.

The two dialectal groups do not seem to differ with respect to their dominance scores, $t(85.1) = -0.608$, $p > 0.05$ [.54], 95% c.i. [−14.9, 7.9], Cohen's $d = -0.13$, but they do in regards to their English vocabulary size scores, $t(82.7) = -3.63$, $p < 0.001$ [0.0004], 95% c.i. [−11.86, −3.47], Cohen's $d = -0.77$. The average BLP score for the Querétaro group is −97.3 (*SD* = 28.3, *range* [−142.2, −18.1]), and the average for the Sonora group is −100.8 (*SD* = 25.5, *range* [−140.1, −41.5]). This confirms that all participants are dominant in Spanish. The average LexTALE score for the Querétaro group is 69.6 (*SD* = 10.8, *range* [53.7, 97.5]), and the average for the Sonora group is 61.9 (*SD* = 8.9, *range* [42.5, 81.25]). Thus, the Queretaroans have, on average, higher vocabulary size scores than the Sonorans, but there is much overlap between the two groups. On average, neither of the two groups are near ceiling (i.e., 90% or higher). In terms of their vocabulary size scores, both groups include a relatively wide range of learners.

2.2. Materials

2.2.1. Identification Experiment

The key data in this study were collected by means of two perception tasks, an identification task and a categorical discrimination task. In the identification task, participants were presented with 96 auditory stimuli consisting of one of four English words: *cheat*, *sheet*, *seat*, and *sit*. A total of 24 different iterations of each of the four words were played in random order to each participant. Listeners were asked to identify each stimulus by indicating, from a closed list of options, the lexical item each auditory stimulus corresponded to. Four options to select from were shown on a computer screen in alphabetical order, from left to right: *cheat*, *sheet*, *seat*, *sit*. The participants are hypothesized to misidentify *cheat* as *sheet*, *sheet* as *cheat*, *seat* as *sit*, and *sit* as *seat*. We have no hypothesis as to whether they would also misidentify *seat* or *sit* as *cheat* or *sheet*.

2.2.2. Discrimination Experiment

An ABX categorical discrimination task was designed to test two key contrasts: *sheet–cheat* (target) and *seat–sit* (control). In an ABX task, listeners hear a triad of auditory tokens (A, B, and X) presented in a sequence within the same trial and, upon hearing all three, they indicate whether the third token (X) matches either the first (A) or the second (B) item in the sequence. There were no "catch" trials in our version of the task, which means that there always was a correct answer. Importantly, all of the stimuli in each triad were acoustically different, including the two matching tokens, as each one of them had been recorded by a different talker. Under such conditions, comparisons cannot be based on acoustic memory, but must be based on phonological or lexical memory (participants are comparing abstract categories, not auditory tokens), which requires participants to access their mental representations to make their decisions. This is the reason why we refer to this task as a *categorical* ABX.

Each participant provided 48 observations to the data set: 24 trials focused on the *seat–sit* contrast (*sea*t–sit–seat [6]; seat–sit–sit [6]; sit–seat–sit [6]; sit–seat–seat [6]), and 24 focused on the sheet–cheat contrast (sheet–cheat–cheat [6]; sheet–cheat–sheet [6]; sheet–sheet–cheat [6]; sheet–cheat–cheat [6]). In 24 of the trials, the matching word was adjacent to the target word—it was in the second position. In other words, the target word was always in the third position of the sequence and, in cases of adjacency, the matching word was in second position. In 24 of the trials, the matching word was not adjacent to the target—it was in the first position. Everything else being equal, matching adjacent categories is expected to be easier than matching non-adjacent one (Best et al. 2001).

2.2.3. Auditory Stimuli

Four native English speakers, all of them women, served as talkers. Their productions were recorded in a sound-treated booth using professional recording equipment: a Shure SM10A head-mounted dynamic microphone and a Sound Devices USBPre2 audio interface connected to a laptop computer. Speech productions were digitized at 44.1 kHz, with 16-bit quantization. Sound files were normalized for intensity.

The talkers were asked to produce the target words by embedding them in a constant carrier phrase, "__ is the word." The materials were presented in random order to avoid any possible systematic effects of list intonation or exhaustion on the same lexical items. Talkers produced all target words four times (4 tokens $\times$ 4 iterations $\times$ 4 talkers = 64 items). One token of each target word per talker was selected (avoiding disfluencies and any extraneous noise) for a total of four target stimuli per lexical item.

2.3. Procedure

Participants completed the tasks individually. In Querétaro, participants were tested in a sound-attenuated booth, while in Sonora they were tested in a quiet library room. Stimuli were presented auditorily over a set of Audio Technica ATH-M50x closed-circumaural headphones connected to a laptop computer running PsychoPy2 (Peirce et al. 2019). Participants responded by pressing a key on a Logitech G512 Lightsync RGB mechanical keyboard. Prior to the completion of the experimental tasks, the first author, a native Spanish speaker from Sonora, provided them with a general description of the tasks and their instructions. This conversation took place in Spanish. Before participating in any of the perceptual tasks, people completed the Bilingual Language Profile questionnaire (Gertken et al. 2014), then the LexTALE (Lemhöfer and Broersma 2012).

For the identification task, participants were instructed to listen to each stimulus, one per trial, and indicate their answer as quickly and accurately as possible by pressing one of four keys on the keyboard (1, 2, 3, or 4). Trials began with a red cross in the center of the screen for 250 ms, which was followed by a screen showing the four response options: *cheat*, *sheet*, *seat*, *sit*. Words were shown in capital letters. Numbers—that is, key codes—were presented in yellow and shown below their corresponding lexical item. Response options were shown for 2500 ms. Auditory stimuli were played 500 ms from the onset of the screen displaying the response options. Participants were allotted 2 s to enter a response. If participants did not provide a response within the allotted time, a new trial began, and the response was left empty.

In the ABX task, participants were asked to listen to all three sounds presented in the trial and only then respond by pressing either number 1 or number 2 on the keyboard to indicate whether they believed the third sound matched the first (1) or the second (2) one in the triad. The words 'first' and 'second' were shown on the computer screen in upper case and accompanied with their matching key codes, 1 or 2. Each trial began with the showing of a red cross in the center of the screen for 1000 ms. The first stimulus of the triad was played at the 1 s mark and was then followed by the second sound of the triad at the 2 s mark. The stimulus onset asynchrony of these two stimuli was thus set at 1 s. The stimulus onset asynchrony between the second and third stimuli was set at 1.5 s. Simultaneously with the playing of the third auditory stimulus in the triad, a screen showing the two response options was shown. Participants had 2 s to introduce their answer. If no answer was entered within this time, a new trial began, and the response was left empty.

2.4. Analysis

All statistical analyses were run in *R*, with packages *tidyverse* (Wickham 2017), *afex* (Singman et al. 2018), and *effsize* (Torchiano 2018). For reproducibility, readers may obtain the *R* scripts and synthetic data frames from either of the authors.

2.4.1. Identification

The analysis of the identification-task data was conducted in two steps. In the first step, we classified what participants heard (the lexical items the talkers had produced) as a function of what they responded (the lexical items the listeners had responded they had heard). This results in a contingency table. The original data set was comprised of 8448 rows, all of them listeners' responses to auditory stimuli. Nevertheless, a number of these observations were excluded from the analysis because the listener did not respond within their allotted time. There was a total of 424 not-responded-to trials, about 5% of the observations. The analysis was then conducted without such trials, with a data set containing 8024 observations.

In a second step, we simply calculated the proportion of times a given participant was accurate versus the times they were inaccurate. In order to prepare the data for the statistical analysis, we ran an arcsine transformation of the proportion-correct scores by participant and condition.

2.4.2. Discrimination

The original data set comprised a total of 4224 rows, 44 (listeners) × 2 (locations) × 48 (responses), of which 288 (6%) were empty, that is, trials that did not contain any information because the participant had failed to respond within the allotted time. An analysis of the participants' responses was then conducted after removing the empty observations from the data set, which results in a data frame comprised of 3936 observations. The analysis counted the proportion of correct responses per listener, per condition (contrast type and adjacency). Then, the accuracy scores (or proportion of correct responses) were arcsine-transformed for the statistical analysis.

3. Results

3.1. Identification

The analysis of the identification data focuses on the proportion of times the auditory stimuli were identified as each of the four lexical items. Table 1 shows the proportion of responses, calculated only for the trials that were responded to, as a function of stimulus played (rows) and response given (column), further broken down by region-of-origin of the participants.

Table 1. Proportion of times each auditorily presented lexical item was identified as being an instance of one of four possible words (*cheat*, *sheet*, *seat*, *sit*), further broken down by region of origin, in Mexico, of the English learners (Sonora, Querétaro).

	Sonora				**Querétaro**			
	Cheat	***Sheet***	***Seat***	***Sit***	***Cheat***	***Sheet***	***Seat***	***Sit***
[tʃ]*eat*	0.51	0.42	-	-	0.83	0.16	-	-
[ʃ]*eet*	0.39	0.57	-	-	0.20	0.73	-	-
s[iː]*t*	-	-	0.32	0.64	-	-	0.42	0.55
s[ɪ]*t*	-	-	0.64	0.32	-	-	0.55	0.40

Note: Rows represent the auditory stimuli played and columns represent the labels displayed on the screen and, thus, the responses available to the participants. Responses below 5% are not shown. Within each participant group, rows add up to 1.

As it may be observed in Table 1, the proportion values suggest that neither *seat* nor *sit* are likely to be categorized as neither *sheet* nor *cheat*. We may conclude that [s] is categorized as being distinct from both [tʃ] and [ʃ], and that this is true for both groups of learners—when a word begins with [s] and ends with [t], only *seat* and *sit* are viable options. Equivalently, when a word begins with a postalveolar obstruent, either [tʃ] or [ʃ] and end with [t], neither *seat* nor *sit* are viable options. We infer that it is reasonable for us to treat the *sheet–cheat* and the *seat–sit* contrasts as separate binary oppositions in our analysis. The scores in Table 1 also suggest that the identification of both *seat* and *sit* lead to a large number of categorization errors, and that both groups of learners are likely to confuse the two words with each other.

We now turn our attention to *cheat* and *sheet*. It appears that the identification patterns of these two words differ in the two groups of learners. In the case of Queretaroans, *cheat* and *sheet* do not appear to be very difficult to identify even in a task that plays these words in the context of each other—accuracy rates are relatively high, with 83% correct responses for *cheat* and 73% correct responses for *sheet*, but they can nevertheless be confused with each other at rates that are not negligible. On the other hand, the Sonorans made many categorization errors for *cheat* and *sheet*. Sonorans' accuracy rates are relatively low, with 51% correct responses for *cheat* and 57% for *sheet*.

Our statistical analysis focuses on accuracy rates. We select only the cells that may be interpreted as displaying 'correct' responses: [tʃ]*eat* identified as *cheat*, [ʃ]*eet* identified as *sheet*, *s*[iː]*t* identified as *seat*, and *s*[ɪ]*t* identified as *sit*. This analysis ignores all other cells. The arcsine-transformed accuracy scores are analyzed with a mixed-design, two-way $2 \times (4)$ ANOVA with Location (Querétaro, Sonora) as a between-subjects factor, and Item (*cheat*, *sheet*, *seat*, *sit*) as a within-subjects factor. The ANOVA yields main effects of Location, $F(1, 86) = 18.2$, $p < 0.0001$, $\eta^2 = 0.10$, and of Item, $F(2.13, 183.4) = 62.5$, $p < 0.0001$, $\eta^2 = 0.26$. It also yields a statistical interaction between the two factors, $F(2.13, 183.4) = 6.4$, $p < 0.05$ [.002], $\eta^2 = 0.03$. The results reveal that, as a group, the Sonorans are more likely to make categorization mistakes than the Queretaroans with this closed lexical set, but this further depends on the lexical item.

To investigate the interaction, we divide the data set into four subsets as a function of Item. The results reveal that, for both *seat* and *sit*, accuracy rates are comparable across the two learner groups: *sit*, $t(83.3) = -1.16$, $p > 0.05$ [0.25], 95% c.i. [−0.287, 0.075], Cohen's $d = -0.247$; *seat*, $t(66.5) = -1.97$, $p > 0.05$ [0.053], 95% c.i. [−0.252, 0.001], Cohen's $d = -0.42$. In other words, both groups of participants are similarly likely to be accurate when identifying these two words. On the other hand, accuracy rates are different across learner groups for both *cheat*, $t(83.9) = -8.08$, $p < 0.0001$, 95% c.i. [−0.501, −0.303], Cohen's $d = -1.722$, and *sheet*, $t(73.02) = -3.53$, $p < 0.001$ [0.0007], 95% c.i. [−0.316, −0.088], Cohen's $d = -0.753$. In both cases, the Queretaroans are more likely to be accurate than the Sonorans.

To summarize, identifying the two members of the *seat–sit* contrast appears to be similarly challenging for both groups of Spanish-speaking learners of Spanish, whereas asking participants to identify the two members of the *sheet–cheat* contrast is more likely to lead to errors for the Sonorans than for the Queretaroans. The results obtained in the identification task suggest the following hypotheses: (i) The Sonorans are just as likely as the Queretaroans to find the *seat–sit* contrast difficult to discriminate, and so we use this contrast as our control condition in the discrimination study; (ii) the Sonorans are likely to find the *sheet–cheat* contrast more difficult to discriminate than the Queretaroans.

3.2. Discrimination

Table 2 shows the untransformed proportion of correct responses by participant group and experimental condition. There are two experimental conditions in our design: (i) the lexical contrast tested in a given trial (*seat–sit*, *sheet–cheat*), and (ii) the adjacency condition between the target word and the matching one. When the matching stimulus is located in the first position in the triad, the matching and the target stimuli are not adjacent (*primacy* condition), whereas when the matching stimulus is found in the second position in the triad the matching and the target stimuli are adjacent (*recency* condition). Everything else being equal, recency trials are predicted to be easier to answer accurately than primacy ones, particularly for challenging contrasts (Best et al. 2001).

Table 2. Proportion of correct responses by learner group (Querétaro, Sonora), as a function of lexical contrast (*seat–sit*, /iː/-/ɪ/; *sheet–cheat*, /ʃ/–/tʃ/) and adjacency condition (primacy, recency).

	Sonora			**Querétaro**		
	Primacy	**Recency**	***M***	**Primacy**	**Recency**	***M***
/ʃ/–/tʃ/	0.48	0.64	0.56	0.61	0.74	0.68
/iː/-/ɪ/	0.79	0.82	0.81	0.80	0.85	0.83

Note: Primacy stands for trials in which target and matching stimuli are not adjacent; recency stands for trials in which target and matching stimuli are adjacent.

Firstly, we submit the arcsine-transformed proportion-correct scores to a mixed-design, three-way $2 \times (4) \times (2)$ ANOVA with Location (Querétaro, Sonora) as a between-subjects factor, and Contrast (*sheet–cheat*, *seat–sit*) and Adjacency (primacy, recency) as within-subjects factors. The ANOVA yields main effects of Location, $F(1, 86) = 5.7$, $p < 0.05$ [.02], $\eta^2 = 0.02$, Contrast, $F(1, 86) = 117.8$, $p < 0.0001$, $\eta^2 = 0.19$, and Adjacency, $F(1, 86) = 14.8$, $p < 0.001$ [.0002], $\eta^2 = 0.05$. Of these effects, the largest one is Contrast, then Adjacency, then Location. Importantly, there are two significant two-way interactions: Contrast by Adjacency, $F(1, 86) = 4.43$, $p < 0.05$ [.04], $\eta^2 = 0.009$, and Contrast by Location, $F(1, 86) = 6$, $p < 0.05$ [.02], $\eta^2 = 0.01$. There is no significant Location by Adjacency interaction and no significant three-way interaction.

The interactions are explored in three steps. Firstly, to explore the Contrast by Adjacency interaction, we average over Location and analyze the effects of adjacency for the two contrasts separately. This analysis pools the data for the two dialectal regions. According to a series of paired-sample *t*-tests, Adjacency does not trigger a significant effect for the *seat–sit* contrast, $t(87) = 2.05$, $p > 0.016$ [.04], 95% c.i. [0.002, 0.138], Cohen's $d = 0.199$, but it does for the *sheet–cheat* one, $t(87) = 3.9$, $p < 0.001$ [0.00019], 95% c.i. [0.084, 0.259], Cohen's $d = 0.396$. People are found to be less accurate in primacy trials than in recency ones, but most obviously so in trials that test the *sheet–cheat* contrast than the other one. Secondly, to explore the Contrast by Location interaction, we average over Adjacency and analyze the effects of Contrast for the two learner groups separately. For both groups, the *sheet–cheat* contrast leads to significantly more response errors than the *seat–sit* contrast (Querétaro: $t(87) = -6.27$, $p < 0.0001$, 95% c.i. [−0.27, −14], Cohen's $d = -0.595$; Sonora: $t(87) = -8.92$, $p < 0.0001$, 95% c.i. [−0.397, −0.252], Cohen's $d = -0.971$), but the effect is larger for the Sonorans than for Queretaroans. Finally, returning once more to the Contrast by Location interaction, the potential effects of Location for the two contrasts are analyzed separately. According to two Welch t-tests, there is no significant effect of Location for the *seat–sit* contrast, $t(172.5) = -0.616$, $p > 0.0125$ [.54], 95% c.i. [−0.108, 0.057], Cohen's $d = -0.093$, whereas the effect is statistically significant for the *sheet–cheat* contrast, $t(173.9) = -3.37$, $p < 0.0125$ [.0009], Cohen's $d = -0.508$.

To summarize, both groups learners find both lexical contrasts relatively difficult to discriminate. Interestingly, the *sheet–cheat* contrast appears to be more challenging than the *seat–sit* contrast. Additionally, both groups are similarly accurate in their discrimination of the *seat–sit* contrast, which we are taking to be our control condition. The most important finding is that, for the *sheet–cheat* contrast, the Sonorans are more likely than the Queretaroans to make discrimination errors. From these analyses, one could infer that the discrimination of the /ʃ/–/tʃ/ contrast is more challenging for the Sonorans than it is for the Queretaroans. Recall, however, that, according to the LexTALE, the Queretaroans in our sample have, on average, a larger English vocabulary than the Sonorans. Thus, the finding could be due, to some extent, to an asymmetry in English vocabulary size rather than to their native dialect phonologies.

To address the possibility that vocabulary size, rather than native phonology, explains these findings, we select a subset of the 22 learners with the lowest LexTALE scores in the Querétaro group and the 22 learners with the highest LexTALE scores in the Sonora sample to form a subset comprising 44 learners, $\frac{1}{2}$ of the sample. In this subset, the vocabulary size of the Sonorans ($M = 68.6$, $SD = 5.9$) is larger than that of the Queretaroans ($M = 61$, $SD = 4.2$), according to a Welch t-test conducted on LexTALE scores, $t(37.8) = 4.89$, $p < 0.0001$, 95% c.i. [4.47, 10.76], Cohen's $d = 1.47$. A mixed-design ANOVA with arcsine-transformed accuracy scores obtained in the discrimination task yields only main effects of Contrast, $F(1, 42) = 76.5$, $p < 0.0001$, $\eta^2 = 0.21$, and Adjacency, $F(1, 42) = 4.47$, $p < 0.05$ [.04], $\eta^2 = 0.04$, and a Contrast by Adjacency interaction, $F(1, 42) = 5.91$, $p < 0.05$ [.02], $\eta^2 = 0.02$, but no other main effects, and no further interactions. Importantly, there are no main effects of Location, $F(1, 42) = 0.023$, $p > 0.05$ [.64], $\eta^2 = 0.002$, and Location does not interact with Contrast, $F(1, 42) = 1.92$, $p > 0.05$ [.17], $\eta^2 = 0.007$. Overall, participants are more accurate in their discrimination of the *seat–sit* contrast ($M = 0.81$, $SD = 0.2$) than of the *sheet–cheat* contrast ($M = 0.59$, $SD = 0.24$). They are also more accurate in recency trials ($M = 0.75$, $SD = 0.23$) than in primacy trials ($M = 0.65$, $SD = 0.25$). Additionally, the effects of adjacency are more robust in the *sheet–cheat* contrast than in the control contrast. In sum, an analysis of a subset of data in which the Sonorans have larger English vocabularies than the Queretaroans fails to reveal any differences between the groups in regard to their discrimination of the *sheet–cheat* contrast (or the *seat–sit* one, for that matter). Apparently, for the Sonorans to match the Queretaroans in their discrimination of the *sheet–cheat* contrast, they must have larger vocabularies than them.

In a final analysis, we explore the potential effects of vocabulary size on the discrimination of the target contrast by means of two linear regression models. These analyses address the following question: Do learners with larger vocabularies show increased sensitivity to the target consonant contrast? To conduct these comparisons, we select one of the experimental conditions—the most "difficult" one, the arcsine-transformed *sheet–cheat* contrast in primacy trials—so that we obtain a single accuracy score per participant, and then correlate this variable with the learners' LexTALE scores. We conduct two regression analyses, one per participant group. Whereas the regression model analyzing the Querétaro data yields a significant finding, $F(1, 42) = 11.2$, $p = 0.002$, adjusted $R^2 = 0.191$, the one analyzing the Sonorans does not, $F(1, 42) = 0.03$, $p < 0.001$, adjusted $R^2 = -0.023$. In other words, the Queretaroans with larger vocabularies seem to be more sensitive to the *sheet–cheat* contrast than the ones with smaller vocabularies, whereas no such relation exists for the Sonorans. According to a series of one-sample t-tests, in this particular experimental condition, the Sonorans (as a group) are not found to have accuracy rates higher than chance (0.5 proportion-correct scores), $t(43) = -0.46$, $p > 0.025$ [0.64], 95% c.i. [0.69, 0.84], while the Queretaroans are, $t(43) = 3.03$, $p < 0.025$ [.002], 95% c.i. [0.84, 1.007].

4. Discussion

4.1. Summary of Findings

The present study reported on two categorization experiments aimed at investigating the acquisition of the English /ʃ/–/tʃ/ (*sheet–cheat*) contrast by two groups of foreign-language learners, both with Spanish as their L1. The first group was recruited in central Mexico (Querétaro), where the local dialect uses [tʃ] but not [ʃ]. The second group was recruited in northwestern Mexico (Sonora), where the local dialect uses both [tʃ] and [ʃ] in free variation. Alongside the target contrast, we investigated the English /i/–/ɪ/ contrast as a control condition.

The findings of a perceptual identification experiment indicated that the learners found the members of the /ʃ/–/tʃ/ contrast difficult to identify when played in the context of each other. On average, 29% of the relevant stimuli were identified incorrectly. Interestingly, the Sonorans (40% mean error rate) made significantly more errors than the Queretaroans (18% mean error rate). The findings of a categorical discrimination experiment demonstrated that the learners found the members of the target contrast hard to discriminate from each other. Overall, the mean error rate for the /ʃ/–/tʃ/ contrast was 38%. Importantly, the groups differ from each other in their discrimination of the /ʃ/–/tʃ/ contrast. In particular, the average error rate of the Sonorans in discrimination of this contrast was as high as 44% while that of the Queretaroans was 32%.

The learner groups differed from each other in their identification and discrimination accuracy of the /ʃ/–/tʃ/ contrast, but not in that of the /i/–/ɪ/ contrast. Taken together, the findings of these two experiments suggest that the Sonorans find the acquisition of the English /ʃ/–/tʃ/ contrast more difficult than the Queretaroans.

4.2. Interpretation and Implications

The experimental evidence suggests that the English /i/–/ɪ/ contrast is particularly difficult for native Spanish speakers who are learning English as a L2 (Escudero and Boersma 2004; Kondaurova and Francis 2008; Morrison 2008, 2009). Studies demonstrate that Spanish-speaking learners of English tend to associate both of the English vowels to their native /i/, perhaps initially merging the three phonetic categories into one. Spanish speakers can certainly overcome this initial obstacle and can acquire this contrast. Nevertheless, even when they do, their representation of the /i/–/ɪ/ distinction is likely to be based on duration, not spectrum; that is, whereas native, monolingual speakers of English distinguish these two vowel categories based on their spectral properties (i.e., first- and second-formant frequencies), native Spanish speakers who have been able to acquire this contrast are likely to base their distinction on a seemingly parasitic correlate, duration (Kondaurova and Francis 2008; Morrison 2008). The results pertaining to both of our participant groups demonstrate that the identification and discrimination of items implementing the /i/–/ɪ/ contrast are challenging for this population.

Let us now focus on an unexpected secondary finding pertaining to the control contrast. It seems that learners were more likely to be accurate in the discrimination task than in the identification one. For the /i/–/ɪ/ contrast condition, the participants were above chance in their discrimination patterns while they were at chance (as a group) when asked to identify *sit* and *seat*. In fact, some participants were more likely to be wrong than right in the identification experiment. One possible interpretation is that our learners have been able to develop separate phonetic categories for /i/ and /ɪ/ while not having learned which word has which sound category. The results of the discrimination task suggest that participants can distinguish the two categories, though neither perfectly nor consistently; those of the identification task, on the other hand, suggest that participants do not associate the phonetic categories with lexical items. Insofar as a phoneme is a phonetic category associated to a particular lexical set (i.e., a category included in a phonolexical representation) (Simonet 2016), one could say that our participants may have developed (fuzzy) separate phonetic, but not *phonemic*, categories for /i/ and /ɪ/. The input they have received may have allowed them to form two phonetic categories, perhaps by means of distributional acoustic learning (Escudero and Williams 2014; Wanrooij et al. 2013); it may not have been sufficient, however, for them to form accurate, detailed phonolexical representations that include those categories. Indeed, experimental evidence suggests that forming phonetic categories in a L2 and associating them with phonolexical representations involve different stages of learning (Amengual 2016; Díaz et al. 2012; Sebastián-Gallés and Díaz 2012).

One factor that may have affected participants' identification patterns in the present study has to do with sound-to-spelling correspondences—recall that participants were asked to identify auditory stimuli in terms of visual ones. In Spanish, the phoneme-to-grapheme correspondence of /i/ consistently matches <i>. Note that the English lexical items chosen for the present study, *seat* and *sit*, had <i> corresponding with /ɪ/, not /i/. It is possible that the learners were variably affected by the Spanish spelling conventions during their identification of *sit*. This, together with their not having developed separate *phonemic* categories for these sounds, could explain the pattern of results. At any rate, the crucial finding pertaining to the /i/–/ɪ/ contrast is that the performance of the two groups of participants in the current study was comparable. We are justified to consider this contrast our control condition.

We now discuss the findings pertaining to the target contrast, /ʃ/–/tʃ/. We hypothesized that the perceptual behavior of our two groups of learners would differ in terms of their categorization of the English /ʃ/–/tʃ/ contrast. This hypothesis was based on the characteristics of the phonology of their native dialect of Spanish. An interesting finding was that both groups of participants seemed to find the English contrast somewhat challenging to acquire. This was not unexpected, as no variety of Spanish has a comparable contrast. The Queretaroans speak a variety of Spanish that uses [tʃ], but not [ʃ], whereas the Sonorans speak a variety that uses both [tʃ] and [ʃ] in free (i.e., not phonologically conditioned) variation. Whereas, as mentioned, all participants found the target contrast somewhat challenging to discriminate, the Queretaroans were found to be, on average, more accurate in their perceptual performance involving this particular contrast than the Sonorans. We propose that the acquisitional obstacles encountered by these two groups of learners, and thus their learning paths, differ on substantial grounds.

Let us first discuss the case of the Queretaroans. As mentioned in the Introduction, most current models of L2 phonological acquisition postulate that learners develop connections between the sound categories of their L2 and those of their L1. One such model is the PAM-L2 (Best and Tyler 2007) and another is the L2LP (Escudero 2005; van Leussen and Escudero 2015). The Queretaroans in our sample may have assimilated both English categories, /tʃ/ and /ʃ/, to the closest category in their L1, /tʃ/. Since Spanish has /tʃ/ but also has voiceless fricatives (/s f h/, just not /ʃ/), one could hypothesize that the acquisitional obstacle encountered by these learners is not insurmountable, provided that these learners extrapolate this aspect of their native system. Speakers of this dialect possess the capacity to represent affricates as being distinct from fricatives. We postulate that the Queretaroans in our sample have assimilated both English /tʃ/ and /ʃ/ to the same native phoneme but, crucially, have done so at different degrees of goodness of fit. Thus, whereas English /tʃ/ may be strongly assimilated to Spanish /tʃ/ (i.e., the match is close to perfect), the interlingual assimilation of English /ʃ/ to Spanish /tʃ/ may be rather poor (i.e., the match is imperfect). In the terminology used in the PAM, this would be an instance of CG assimilation. In cases of CG assimilation, the perceptual discrimination of the two members of the L2 contrast is expected to range from moderate to good. The Queretaroans' discrimination of the English /ʃ/–/tʃ/ contrast is indeed moderate. In sum, we believe that the obstacle the Queretaroans face when learning the English /ʃ/–/tʃ/ contrast is having to develop a new phonetic category, separating from an initial stage in which the two sounds as categorized as instances of the same sound. Postulating that the Queretaroans assimilate both English /tʃ/ and /ʃ/ to the same native phoneme, but at different degrees of fit, explains both the presence of the obstacle and its size (i.e., moderate).

Let us now discuss the case of the Sonorans. In Sonora, as well as in other northwestern Mexican states, the local Spanish dialect uses both [ʃ] and [tʃ] in free variation. Crucially, these phonetic variants are not in complementary distribution, but occupy the same segmental slots in the same lexical items. The obstacle for these particular learners lies in the fact that they must unlearn the phonological mapping patterns of their native dialect—the mapping between phonetic categories (or allophones) and phonolexical representations (or phonemes)—before they can learn those of English. Assuming that in the first stage of L2 acquisition learners transfer their L1 competence into their L2 system (Escudero 2005; van Leussen and Escudero 2015), Sonorans begin their development at a stage in which [ʃ] and [tʃ] are equivalent at some level of representation. Therefore, they must first unlearn that [ʃ] and [tʃ] are variants of the same phoneme before they can learn that these two sounds are contrastive in English and, thus, associated with different lexical sets. The acquisitional obstacle may be formalized in two ways.

The first way in which the acquisitional obstacle encountered by the Sonorans may be formalized makes use of the same theoretical constructs we have used to explain the behavior of the Queretaroans, interlingual phonetic category assimilations (Best and Tyler 2007; Escudero 2005). It appears that the lexical distribution of phonetic categories determines to some extent peoples' perceptual behavior. In particular, phonetic categories in complementary distribution are less likely to be perceived as being distinct from each other than categories that are contrastive in the lexicon (Johnson and Babel 2010). If two sound categories are in free variation, it is even more likely that they will be perceived as being perceptually very similar to each other. Sonorans, therefore, are likely to perceive [ʃ] and [tʃ] as being perceptually more similar (to each other) than speakers of other Spanish dialects. If we extrapolate this to interlingual interactions in L2 acquisition, we postulate that Sonorans have assimilated both English /tʃ/ and /ʃ/ to the same native phoneme, and that both English categories are optimal matches to this native phoneme. The situation for the Sonorans could be one of SC assimilation (Best et al. 2001; Best and Tyler 2007), a *new scenario* (Escudero 2005). In cases of SC assimilation, discriminability is predicted to be very poor. Indeed, the discrimination of the English /ʃ/–/tʃ/ contrast by the Sonorans is very poor.

The second way in which the acquisitional obstacle encountered by the Sonorans may be formalized makes use of the concept of *allophonic split* (Barrios et al. 2016b; Eckman et al. 2001). This formalization does not depend on interlingual perceptual assimilations between phonetic categories but makes use of the concept of mapping between surface and underlying representations. Surface allophones that find themselves in either free variation or complementary distribution are mapped onto a single underlying segment, a phoneme. A Sonoran learner of L2 English would need to establish new mappings between familiar phones (Barrios et al. 2016b). The phonological competence of Sonoran Spanish speakers includes both [ʃ] and [tʃ], but, since the two sounds are mapped onto the same phoneme, learning the English /ʃ/–/tʃ/ contrast would require creating a new underlying representation and mapping the two surface allophones to separate phonemes. Eckman et al. (2001) hypothesize that this scenario is particularly challenging for L2 learners, but Barrios et al. (2016b) did not find evidence to support this claim. Insofar as we can conceptualize the learning scenario of Sonoran learners as a case of allophonic split, our data are fully in line with Eckman's hypothesis. Interestingly, our data suggest that cases of allophonic split (Sonora) are more challenging to overcome than cases of new category formation (Querétaro).

Since Sonoran learners possess a phonological system that includes both [ʃ] and [tʃ], one could have hypothesized that acquiring the English /ʃ/–/tʃ/ contrast would be particularly easy in their case. Obviously, this is not what our data suggest. Our data suggest that, in addition to existing phonetic categories, lexico-distributional patterns (i.e., the patterns of lexical distribution of phones, which determine contrastivity, among other things) determine, to some extent, the significance of acquisitional obstacles.

Recall that our two participant samples differed, not only in their region of origin, but also in the sizes of their English vocabularies. We found that, on average, our Queretaroans had slightly larger vocabularies in English than our Sonorans. One could attribute the difference between the two participant groups in terms of their perceptual behavior with respect to the /ʃ/–/tʃ/ contrast to their overall proficiency in English. The Queretaroans may have been more accurate than the Sonorans in their discrimination of the /ʃ/–/tʃ/ contrast because they may be generally more proficient in English than the Sonorans. While this is certainly possible, we would like to highlight the following. Firstly, along with our main data set, we compared the behavior of Sonorans and Queretaroans in a controlled subset. For this subset, we selected 10 Sonorans and 10 Queretaroans so that the Sonorans had larger vocabularies than the Queretaroans. In this subset, participants' perceptual behaviors with respect to the /ʃ/–/tʃ/ contrast were found to be indistinguishable. Sonorans with larger vocabularies discriminate the /ʃ/–/tʃ/ contrast just as poorly as Queretaroans with smaller English vocabularies—thus, vocabulary size is not the only determinant of /ʃ/–/tʃ/ discrimination. Secondly, the control contrast condition, /i/–/ɪ/, led to comparable behavior across the two groups—whereas vocabulary sizes may differ between groups, their overall state of phonological development may not. Thirdly, Queretaroans, but not Sonorans, seem to become "better" at discriminating the /ʃ/–/tʃ/ contrast as their vocabulary increases. These observations suggest that the difference between Sonorans and Queretaroans reported in the present study is indicative of a larger issue, such as a difference in their native phonologies, not simply a difference in overall English competency. Nevertheless, only future research can resolve this conundrum.

5. Conclusions

The present study investigated the significance of two types of acquisitional obstacles in L2 phonology. The study reported on the identification and discrimination of the English /ʃ/–/tʃ/ (*sheet–cheat*) contrast by two groups of learners whose L1 is Spanish. The first group was recruited in central Mexico, where the local dialect uses [tʃ] but not [ʃ]. To learn the English /ʃ/–/tʃ/ contrast, speakers of this Spanish variety must learn a new phonetic category, [ʃ]. The second group was recruited in northwestern Mexico, where the local dialect uses both [tʃ] and [ʃ] in free variation. Since both obstruents are variants of the same phoneme in this variety, to learn the English /ʃ/–/tʃ/ contrast, speakers of this dialect must develop new mapping between familiar phonetic categories and underlying (contrastive) representations. The study found that the acquisitional obstacle encountered by speakers of the northwestern Mexican variety of Spanish is of a larger magnitude than the one encountered by speakers from central Mexico. Native dialect phonology is a powerful determinant of L2 phonology learning paths.

Author Contributions: The research study reported here is part of M.L.V.'s doctoral dissertation, supervised by M.S. M.L.V. and M.S. participated in the design and planning of the experiments, and both shared in the analysis of the data. M.L.V. collected all of the data, and M.S. drafted the manuscript. All authors have read and agree to the published version of the manuscript.

Funding: This research was funded by the College of Humanities of the University of Arizona (Graduate Student Research Grant, 2019) and by a Research and Project (ReaP, 2019) grant provided by the Graduate and Professional Student Council of the University of Arizona.

Conflicts of Interest: The authors declare no conflict of interest.

References

Alessi Molina, María T., and Ana Luisa Torres Díaz. 1994. Aspectos fonéticos del habla sonorense. In *Estudios de Lingüística y Sociolingüística*. Edited by Gerardo López Cruz and José Luis Moctezuma Zamarrón. Hermosillo: Universidad de Sonora, pp. 285–92.

Amastae, Jon. 1996. *Variación y Cambio en el Español de Ciudad Juárez, México*. Ciudad Juárez: Universidad Autónoma de Ciudad Juárez.

Amengual, Mark. 2016. The perception of language-specific phonetic categories does not guarantee accurate phonological representations in the lexicon of early bilinguals. *Applied Psycholinguistics* 37: 1221–51. [CrossRef]

Baker, Wendy, and Laura Catharine Smith. 2010. The impact of L2 dialect on learning French vowels: Native English speakers learning Quebecois and European French. *Canadian Modern Language Review* 66: 711–38. [CrossRef]

Barrios, Shannon L., Anna M. Namyst, Ellen F. Lau, Naomi H. Feldman, and William J. Idsardi. 2016a. Establishing new mappings between familiar phones: Neural and behavioral evidence for early automatic processing of nonnative contrasts. *Frontiers in Psychology* 7: 1–16. [CrossRef] [PubMed]

Barrios, Shannon L., Nan Jiang, and William J. Idsardi. 2016b. Similarity in L2 Phonology: Evidence from L1 Spanish late-learners' perception and lexical representation of English vowel contrasts. *Second Language Research* 32: 367–95. [CrossRef]

Best, Catherine T., and Michael D. Tyler. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*. Edited by Ocke-Schwen Bohn and Murray M. Munro. Amsterdam: John Benjamins, pp. 13–34.

Best, Catherine T., Gerald W. McRoberts, and Nomathemba M. Sithole. 1988. Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance* 14: 345–60. [CrossRef] [PubMed]

Best, Catherine T., Gerald W. McRoberts, and Elizabeth Goodell. 2001. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America* 109: 775–94. [CrossRef]

Bohn, Ocke-Schwen. 2017. Cross-language and second language speech perception. In *The Handbook of Psycholinguistics*. Edited by Eva M. Fernández and Helen Smit Cairns. Hoboken: John Wiley & Sons, pp. 213–39.

Broselow, Ellen, and Yoonjung Kang. 2013. Phonology and speech. In *The Cambridge Handbook of Second Language Acquisition*. Edited by Julia Herschensohn and Martha Young-Scholten. Cambridge: Cambridge University Press, pp. 529–54.

Brown, Dolores. 1989. El habla juvenil de Sonora, México: La fonética de 32 jóvenes. *Nueva Revista de Filología Hispánica* 37: 43–82. [CrossRef]

Carreón Serna, Yadira Aimé. 2007. La Lenición en el Español del Norte de México y Otras Lenguas Romances: Una Aproximación a la Variación Sociolingüística y un Modelo Formal Basado en la Teoría de la Optimidad. Ph.D. dissertation, University of California, Santa Barbara, CA, USA, September.

Casillas, Joseph V. 2015. Production and perception of the /i/-/ɪ/ vowel contrast: The case of L2 dominant early learners of English. *Phonetica* 72: 182–205. [CrossRef]

Casillas, Joseph V., and Miquel Simonet. 2016. Production and perception of the English /æ/-/ɑ/ contrast in switched-dominance speakers. *Second Language Research* 32: 171–95. [CrossRef]

Chang, Charles B. 2019. The phonetics of second language learning and bilingualism. In *The Routledge Handbook of Phonetics*. Edited by William F. Katz and Peter F. Assmann. New York: Routledge, pp. 427–47.

Chládková, Kateřina, and Václav Jonáš Podlipský. 2011. Native dialect matters: Perceptual assimilation of Dutch vowels by Czech listeners. *Journal of the Acoustical Society of America* 130: EL186–EL192. [CrossRef] [PubMed]

Colantoni, Laura, Jeffrey Steele, and Paola Escudero. 2015. *Second Language Speech: Theory and Practice*. Cambridge: Cambridge University Press. [CrossRef]

Davidson, Lisa. 2017. Cross-language speech perception and production. In *Oxford Bibliographies in Linguistics*. Edited by Mark Aronoff. Oxford: Oxford University Press. [CrossRef]

Díaz, Begoña, Holger Mitterer, Mirjam Broersma, and Núria Sebastián-Gallés. 2012. Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access. *Learning and Individual Differences* 22: 680–89. [CrossRef]

Eckman, Fred R. 2012. Second language phonology. In *The Routledge Handbook of Second Language Acquisition*. Edited by Susan M. Gass and Alison Mackey. New York: Routledge, pp. 91–105.

Eckman, Fred R., Abdullah Elreyes, and Gregory K. Iverson. 2001. Allophonic splits in L2 phonology: The question of learnability. *International Journal of English Studies* 1: 21–51.

Escudero, Paola. 2005. *Linguistic Perception and Second Language Acquisition: Explaining the Attainment of Optimal Phonological Categorization*. Ph.D. dissertation, Utrecht University, Utrecht, The Netherlands. LOT Dissertation Series 113; Amsterdam: Landelijke Onderzoekschool Taalwetenschap.

Escudero, Paola, and Paul Boersma. 2004. Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language* 26: 551–85. [CrossRef]

Escudero, Paola, and Kateřina Chládková. 2010. Spanish listeners' perception of American and Southern British English vowels. *Journal of the Acoustical Society of America* 128: EL254–EL259. [CrossRef]

Escudero, Paola, and Daniel Williams. 2014. Distributional learning has immediate and long-lasting effects. *Cognition* 133: 408–13. [CrossRef]

Escudero, Paola, Ellen Simon, and Holger Mitterer. 2012. The perception of English front vowels by North Holland and Flemish listeners: Acoustic similarity predicts and explains cross-linguistic and L2 perception. *Journal of Phonetics* 40: 280–88. [CrossRef]

Flege, James E. 1995. Second language speech learning: Theory, findings, and problems. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Edited by Winifred Strange. York: York Press, pp. 229–73.

Flege, James E., and Ocke-Schwen Bohn. 1989. An instrumental study of vowel reduction and stress placement in Spanish-accented English. *Studies in Second Language Acquisition* 11: 35–62. [CrossRef]

Flege, James E., Murray J. Munro, and Robert Allen Fox. 1994. Auditory and categorical effects on cross-language vowel perception. *Journal of the Acoustical Society of America* 95: 3623–41. [CrossRef]

Flege, James E., Ocke-Schwen Bohn, and Sunyoung Jang. 1997. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics* 25: 437–70. [CrossRef]

Gertken, Libby M., Mark Amengual, and David Birdsong. 2014. Assessing language dominance with the Bilingual Language Profile. In *Measuring L2 Proficiency: Perspectives from SLA*. Edited by Pascale Leclerq, Amanda Edmonds and Heather Hilton. Bristol: Multilingual Matters, pp. 208–25.

Hualde, José Ignacio. 2005. *The Sounds of Spanish*. Cambridge: Cambridge University Press.

Jaramillo, June A., and Garland D. Bills. 1982. The phoneme /ch/ in the Spanish of Tomé, New Mexico. In *Bilingualism and Language Contact: Spanish, English, and Native American Languages*. Edited by Florence Barkin, Elizabeth A. Brandt and Jacob Ornstein-Galicia. New York: Teachers College Press, Columbia University, pp. 154–65.

Johnson, Keith, and Molly Babel. 2010. On the perceptual basis of distinctive features: Evidence from the perception of fricatives by Dutch and English speakers. *Journal of Phonetics* 38: 127–36. [CrossRef]

Kondaurova, Maria V., and Alexander L. Francis. 2008. The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *Journal of the Acoustical Society of America* 124: 3959–71. [CrossRef]

Lado, Robert. 1957. *Linguistics across Cultures: Applied Linguistics for Language Teachers*. Ann Arbor: University of Michigan Press.

Lemhöfer, Kristin, and Mirjam Broersma. 2012. Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods* 44: 325–43. [CrossRef]

Lope Blanch, Juan M. 1990. *Atlas Lingüístico de México*. 6 vols. México City: El Colegio de México-UNAM-FCE.

López Velarde, Mariela, and Miquel Simonet. 2019. Spoken word recognition and shesheo in Northwestern Mexico. In *Recent Advances in the Study of Spanish Sociophonetic Perception*. Edited by Whitney Chappell. Amsterdam: John Benjamins, pp. 266–85.

Martín Butragueño, Pedro. 2009. *Regularidad y Excepcionalidad del Cambio Lingüístico: El Caso de (tʃ) en la Geografía Fónica de México*. México City: El Colegio de México.

Martín Butragueño, Pedro. 2011. Dialectología de nuevos mundos: Una lectura variacionista del Atlas Lingüístico de México. In *Selected Proceedings of the 13th Hispanic Linguistics Symposium*. Edited by Luis A. Ortiz-López. Somerville: Cascadilla Proceedings Project, pp. 17–32.

Martín Butragueño, Pedro. 2014. La división dialectal del español mexicano. In *Historia Sociolingüística de México*. Edited by Pedro Martín. México, City: El Colegio de México, pp. 1353–408.

Mayr, Robert, and Paola Escudero. 2010. Explaining individual variation in L2 perception: Rounded vowels in English learners of German. *Bilingualism: Language and Cognition* 13: 279–97. [CrossRef]

Méndez, Luis Alberto. 2017. The variant [ʃ] in the Spanish of Ciudad Juárez. *Borealis—An International Journal of Hispanic Linguistics* 6: 243–60. [CrossRef]

Moreno de Alba, José G. 1994. *La Pronunciación del Español en México*. México City: El Colegio de México.

Morrison, Geoffrey Stewart. 2008. L1-Spanish speakers' acquisition of the English /i/-/ɪ/ contrast: Duration-based perception is not the initial developmental stage. *Language and Speech* 51: 285–315. [CrossRef]

Morrison, Geoffrey Stewart. 2009. L1-Spanish speakers' acquisition of the English/i/—/ɪ/ contrast II: Perception of vowel inherent spectral change. *Language and Speech* 52: 437–62. [CrossRef] [PubMed]

Peirce, Jonathan, Jeremy R. Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman, and Jonas Kristoffer Lindeløv. 2019. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods* 51: 195–203. [CrossRef] [PubMed]

Piske, Thorsten, Ian R. A. MacKay, and James E. Flege. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29: 191–215. [CrossRef]

Sebastián-Gallés, Núria, and Begoña Díaz. 2012. First and second language speech perception: Graded learning. *Language Learning* 62: 131–47. [CrossRef]

Serrano Morales, Julio C. 2000. Contacto dialectal (¿y cambio lingüístico?) en español: El caso de la /tʃ/ sonorense. In *Estructuras en Contexto: Estudios de Variación Lingüística*. Edited by Pedro Martín Butragueño. México City: El Colegio de México, pp. 45–59.

Serrano Morales, Julio C. 2009. Existe el noroeste mexicano como zona dialectal. In *Lengua, Literatura y Región*. Edited by Everardo Mendoza Guerrero, Maritza López Berríos and Ilda Elizabeth Moreno Rojas. Culiacán: Universidad Autónoma de Sinaloa.

Simonet, Miquel. 2016. The phonetics and phonology of bilingualism. In *Oxford Handbooks Online*. Oxford: Oxford University Press. [CrossRef]

Singman, Henrik, Ben Bolker, Jake Westfall, and Frederik Aust. 2018. Afex: Analysis of Factorial Experiments (Version 0.22-1). Computer Software. Available online: https://CRAN.R-project.org/package=afex (accessed on 4 June 2020).

Torchiano, Marco. 2018. Effsize: Efficient Effect Size Computation (Version 0.7.4). Computer Software. Available online: https://CRAN.R-project.org/package=effsize (accessed on 4 June 2020).

Van Leussen, Jan-Willem, and Paola Escudero. 2015. Learning to perceive and recognize a second language: The L2LP model revised. *Frontiers in Psychology* 6: 1–12. [CrossRef] [PubMed]

Wanrooij, Karin, Paola Escudero, and Maartje E.J. Raijmakers. 2013. What do listeners learn from exposure to a vowel distribution? An analysis of listening strategies in distributional learning. *Journal of Phonetics* 41: 307–19. [CrossRef]

Wickham, Hadley. 2017. Tidyverse: Easily Install and Load the "Tidyverse" (Version 1.2.1). Computer Software. Available online: https://CRAN.R-project.org/package=tidyverse (accessed on 4 June 2020).

languages

MDPI

Article

The Contributions of Crosslinguistic Influence and Individual Differences to Nonnative Speech Perception

Charles B. Chang [1,*] **and Sungmi Kwon** [2]

1 Department of Linguistics, Boston University, Boston, MA 02215, USA
2 Department of Korean Language and Literature, Pukyong National University, Busan 48513, Korea; smkwon@pknu.ac.kr
* Correspondence: cc@bu.edu

Received: 1 September 2020; Accepted: 29 October 2020; Published: 31 October 2020

Abstract: Perception of a nonnative language (L2) is known to be affected by crosslinguistic transfer from a listener's native language (L1), but the relative importance of L1 transfer vis-a-vis individual learner differences remains unclear. This study explored the hypothesis that the nature of L1 transfer changes as learners gain experience with the L2, such that individual differences are more influential at earlier stages of learning and L1 transfer is more influential at later stages of learning. To test this hypothesis, novice L2 learners of Korean from diverse L1 backgrounds were examined in a pretest-posttest design with respect to their perceptual acquisition of novel L2 consonant contrasts (the three-way Korean laryngeal contrast among lenis, fortis, and aspirated plosives) and vowel contrasts (/o/-/ʌ/, /u/-/ɨ/). Whereas pretest performance showed little evidence of L1 effects, posttest performance showed significant L1 transfer. Furthermore, pretest performance did not predict posttest performance. These findings support the view that L1 knowledge influences L2 perception dynamically, according to the amount of L2 knowledge available to learners at that time. That is, both individual differences and L1 knowledge play a role in L2 perception, but to different degrees over the course of L2 development.

Keywords: second language acquisition; perceptual learning; individual differences; phonetic sensitivity; crosslinguistic influence; Korean; laryngeal contrast; voice onset time; vowel inventory

---

1. Introduction

Whereas infants are described as "citizens of the world" (Werker and Tees 1984) when it comes to perceiving the sounds of a nonnative language (L2), adults are known to be comparatively poor at L2 speech perception. This disparity follows from a process of perceptual specialization for the native language (L1), which begins well before the end of the first year of life and results in an apparent decline in perception of L2 sound contrasts (Kuhl et al. 1992; Polka and Werker 1994; Werker and Pegg 1992). Although the characterization of this process varies throughout the literature, in recent years researchers have converged upon the view that specialization for the L1 does not involve loss of perceptual ability per se, but rather shifts in attention, which may be attributed to lexical and/or phonemic development (Werker 1994, 1995; Werker and Curtin 2005) or to the impetus for perceptual routines to be automatic and robust to adverse conditions (Strange 2011).

Under the view that mature L1 speakers maintain access to the perceptual and cognitive abilities undergirding L1 phonological development (e.g., Flege 1995), it becomes relevant to ask how, in L2 learning, these underlying abilities interact with the consequences of L1 specialization, especially in light of individual variation in these abilities. That is, how are the processes and outcomes of L2 acquisition affected by language-specific properties of a learner's L1 vs. the personally-specific abilities of the learner? The effect of the L1 has been discussed extensively in terms of TRANSFER or crosslinguistic influence (CLI) of aspects of L1 knowledge and/or experience (Major 2008; Odlin 1989) and plays a central role in theories of L2 speech learning (see Section 1.1). However, the effect of the learner (i.e., INDIVIDUAL DIFFERENCES among speakers who may share the same L1; Dörnyei 2006), long a focus of research in the interdisciplinary field of second language acquisition (SLA), has only recently begun to be considered systematically in the study of L2 perceptual learning (see Section 1.2).

The study reported in this article is an attempt to explore the interaction of L1 transfer and individual differences in L2 perceptual learning of Korean phonological contrasts. In what follows, we review the literature on transfer and individual differences in L2 learning, describe the target Korean contrasts (the three-way laryngeal contrast among lenis, fortis, and aspirated voiceless stops and two vowel contrasts between rounded and unrounded back vowels), and motivate a TRANSFER RAMP-UP HYPOTHESIS predicting a delayed onset of L1 transfer effects in L2 perceptual development.

1.1. Crosslinguistic Influence in Speech Learning

That L2 speech learning may be influenced by previously acquired linguistic knowledge is well-established in the field of L2 speech and in SLA more broadly. Major theories of nonnative speech perception, such as the Perceptual Assimilation Model (PAM; Best 1994, 1995) and its extension to L2 learners, PAM-L2 (Best and Tyler 2007), the Native Language Magnet theory (NLM; Kuhl and Iverson 1995; Kuhl et al. 1992, 2008), the Speech Learning Model (SLM; Flege 1995), and the Automatic Selective Perception framework (ASP; Strange 2011), all incorporate a role for a listener's L1, although they vary in terms of how L1 influence is conceptualized. For example, NLM emphasizes the role of developing L1 category prototypes, whereas PAM(-L2) focuses on the ways in which L2 contrasts may be perceptually assimilated to L1 contrasts, which have consequences for both discrimination and identification of L2 sounds (e.g., Tyler et al. 2014).

Despite their differences, these theories of nonnative speech perception are similar in providing an account for ostensibly negative effects of L1 development in perception of an L2. Although L1 experience is not always a handicap in L2 perception (Bohn and Best 2012; Chang 2016, 2018), it often appears that listeners with an already-developed L1 are at a disadvantage compared to listeners who are less far along in L1 development. For example, English-speaking adults were found to discriminate contrasts of Thompson Salish poorly compared to English-learning infants (Werker and Tees 1984). Such disparities have been interpreted in terms of a decline in sensitivity to nonnative contrasts during L1 development (e.g., Kuhl et al. 1992; Werker and Pegg 1992). It remains unclear, however, whether the disadvantage for adult listeners in L2 perception has to do with reduced phonetic sensitivity or "negative transfer" of L1 structures such as phonological categories (or both).

Indeed, a fundamental component of many theories of L2 speech is the misleading influence of L1 categories. For example, the SLM posits that the most difficult L2 sounds to acquire accurately in the long term are not "new" sounds, but rather "similar" sounds, which resemble the sounds of the L1. This is because a "similar" sound, unlike a "new" sound, gets perceptually linked with an L1 sound, which inhibits the formation of a distinct representation for the L2 sound. According to the SLM, L1 influence (in the form of crosslinguistic perceptual linkage) takes place at the level of the position-specific allophone, but there is reason to believe that the phoneme, rather than the allophone, may be the structure most liable to exert L1 influence. In particular, some findings suggest that access to L1 allophones, which are typically below the level of speaker consciousness, may depend on factors such as the level of acquisition and the type of orthography (Bergier 2014; Eckman et al. 2001; Vokic 2010).

Besides formalizing the influence of L1 categories, theories such as the SLM also posit that "mechanisms and processes used in learning the L1 sound system ... remain intact over the life span, and can be applied to L2 learning" (Flege 1995, p. 239), suggesting that the basic perceptual abilities supporting L1 acquisition are maintained in adulthood. This view is consistent with findings showing little effect of L1 specialization in certain perceptual tasks with nonnative speech. For instance, although [s] and [ʃ] are phonemic in English and allophonic in Dutch, L1 Dutch adults were just as good as L1 English adults at discriminating the two sounds in English nonce items; however, L1 Dutch adults also rated the sounds as more similar than did L1 English adults (Johnson and Babel 2010). This type of result is at odds with the view that adult listeners necessarily have reduced sensitivity to L2 contrasts, and further suggests that any effect of the L1 sound system is closely related to task type (in particular, the levels of processing tapped by the task; cf. Díaz et al. 2012).

If adults are capable of perceiving nonnative contrasts, this still leaves the question of why they tend to perform worse than infants in L2 perception. Strange's (2011) ASP framework addresses this question with the construct of selective perception routines (SPRs). According to ASP, L1 acquisition involves developing SPRs specialized for the target L1. These SPRs are selective in the sense that they target only those acoustic cues that are relevant for distinguishing L1 phonemes, which allows L1 perception to become automatic and robust in adverse conditions. Crucially, a language-universal mode of perception never disappears, but high task demands may block access to this mode, encouraging instead a default to L1 SPRs. In ASP, then, what distinguishes adult and infant listeners is not reduced sensitivity, but shifted attention. As in L1 acquisition, ASP posits that L2 learning involves developing SPRs that weight and direct attention to cues in a manner specialized for the target L2. This implies that L2 learning generally leads to a (positive) change in L2 perception, which has indeed been found (e.g., Kilman et al. 2014; Levy and Strange 2008; cf. Holliday 2016).

Insofar as better L2 perception reflects less L1 transfer, the finding of higher L2 perceptual accuracy in more advanced learners supports the view that L1 transfer decreases over the course of L2 learning. This view is at the heart of a different theory of L2 phonological development, the Ontogeny Phylogeny Model (OPM; Major 2001). OPM differs from other theories of L2 speech in addressing the time course of all three factors that play a role in general SLA theories: L1, L2, and universal components (cf. Universal Grammar; White 2003). The central claim is that, within a learner's developing L2 system, L1 components decrease over time, while L2 components increase; meanwhile, universal components show an inverse U-shaped pattern, first increasing and then decreasing. This theory also formalizes aspects of intra-speaker variation by describing the influence of style/register, crosslinguistic similarity, and markedness. However, inter-speaker variation is not addressed in this model (or in the others discussed above), and is the subject of a different literature.

1.2. *Individual Differences in Speech Learning*

Interest in individual differences (IDs) has existed for decades, going back to the 1970s when SLA researchers started examining the predictive power of various personally-specific properties of the learner (for reviews, see Dörnyei 2006; Dörnyei and Skehan 2003). These include (foreign) language (learning) aptitude, general intelligence or cognitive ability (for a recent review, see Bowles et al. 2016), personality (e.g., Dewaele and Furnham 2000; Verhoeven and Vermeer 2002), and musicality (e.g., Milovanov et al. 2008, 2010), all of which are understood to be complex constructs consisting of several subconstructs (e.g., Robinson 2001). For example, language aptitude may consist of variables such as phonemic coding ability, grammatical sensitivity, inductive learning ability, and rote learning ability. Based on Carroll (1981), Skehan (1989, 1991, 1998, 2002) attempted to explain three abilities (namely, auditory, linguistic, and memory-based), while others (e.g., Jilka et al. 2008) have conceptualized of IDs in terms of a "talent" dimension.

In regard to L2 perception specifically, IDs may reflect a construct that has been called PHONETIC SENSITIVITY in the L2 speech literature (e.g., Kwon 2013; Munro 2008). Note that it is phonetic, as opposed to psychoacoustic, sensitivity that is of interest here due to the neurolinguistic evidence of a "speech-specific origin of individual variability in L2 phonetic mastery" (Díaz et al. 2008, p. 16083). According to Piske (2008), phonetic sensitivity can be distinguished from phonetic or phonological awareness in not necessarily referring to a conscious level of speech processing, and we use this term in a similar sense, to refer generally to sensitivity to properties of the speech signal that may serve to cue a linguistic contrast (regardless of whether or not the contrast exists in the listener's L1). Under the view that the ability to perceive nonnative contrasts does not go away during L1 development, but may be masked by attentional shifts associated with L1 specialization, it becomes relevant to ask whether

IDs in L2 perceptual learning can be traced to variation in learners' (remaining) phonetic sensitivity to nonnative contrasts. That is, might the degree of (non)commitment to perceptual cues relevant for the L1 constitute an "aptitude" for L2 perceptual learning? If not, then what does?

These questions have begun to be explored in a number of studies examining IDs in L2 perception and/or production. For example, research on L1 English speakers trained on Mandarin tonal contrasts found that IDs in learning outcomes were related to learners' pre-training inclination to attend to the most informative cue for Mandarin tones (i.e., pitch direction) as well as to the interaction between their basic pitch perception abilities and the type of training they underwent (Chandrasekaran et al. 2010; Perrachione et al. 2011). Effects of preexisting variation in cue weighting, which can be arbitrary (i.e., not determined by cue informativeness; Idemaru et al. 2012), were also observed in L1 Spanish learners of Dutch and L1 Korean learners of English (Wanrooij et al. 2013; Schertz et al. 2016); however, IDs in L2 perception were not predicted by learners' cue use in L2 production (Schertz et al. 2015). Other studies have linked IDs to variation in compactness and location of L1 categories (Kartushina and Frauenfelder 2013, 2014), self-perception of one's own L2 speech (Baker and Trofimovich 2006), formation and structure of L2 representations (Golestani and Zatorre 2009; Hattori and Iverson 2009), L2 mispronunciation detection (Hanulíková et al. 2012), L2 working memory (Darcy et al. 2015), and different neural correlates (Raizada et al. 2010; Sebastián-Gallés et al. 2012; for a recent review, see Myers 2014).

Some of the work on IDs has focused on the specific target language in the present study—Korean. As reported for learners of other L2s, learners of Korean were found to evince wide variation in L2 perception, both in discrimination of L2 contrasts prior to extensive L2 exposure as well as in identification of L2 categories following weeks of L2 learning; however, pre-learning performance (i.e., phonetic sensitivity to the target L2 contrasts) showed no, or only a weak, correlation with post-learning performance (Jung and Kwon 2010; Kwon 2013). Notably, these results were found with learners representing two or more L1 backgrounds, including typologically diverse and genetically unrelated language families (e.g., Austronesian, Finno-Ugric, Indo-European, Sino-Tibetan). The majority of the data on Korean L2 speech, however, comes from studies focusing on L1 speakers of English and/or Mandarin (e.g., Chang 2010, 2012; Holliday 2014, 2015; Schmidt 2007). These studies show, on the one hand, many dominant patterns of response in perceptual tasks such as crosslinguistic classification (e.g., L1 Mandarin learners tended to classify Korean lenis and aspirated stops as the same Mandarin category) yet, on the other hand, substantial IDs, especially in production (e.g., L1 English learners produced Korean stop contrasts in eight different ways). In the present study, we investigate the interaction of IDs with L1 transfer, focusing on Korean as the target L2. Next, we consider the specific target contrasts, each of which may pose difficulty for L2 learners.

1.3. Korean Phonological Contrasts

1.3.1. Stop Laryngeal Contrasts in Korean

Korean is known for a typologically rare laryngeal contrast among three series of voiceless stops: LENIS, FORTIS, and ASPIRATED. This three-way laryngeal contrast occurs at four places of articulation in all (bilabial, denti-alveolar, and velar stops, as well as alveolo-palatal affricates) and has been described in terms of several different acoustic dimensions, including voice onset time (VOT), closure duration, and properties of the following vowel such as onset fundamental frequency (f_0), voice quality, vowel duration, intensity buildup, and formant trajectories (Cho et al. 2002; Kagaya 1974; Park 2002a, 2002b; for a recent review, see Chang 2013).

The most widely-studied acoustic cues to the stop laryngeal contrasts in utterance-initial position are VOT and f_0. Traditionally, fortis, lenis, and aspirated stops are described as showing, respectively, very short, medium-lag, and very long VOTs, as well as high, low, and very high onset f_0 values. However, studies have noted a recent diachronic shift in the phonetic implementation of the three laryngeal categories, which is especially evident in Seoul Korean and in female speakers (Kang 2014; Oh 2011; Silva 2006a, 2006b). In particular, the VOT of lenis stops has lengthened while that of aspirated stops has shortened, leading to highly overlapping VOT distributions; these developments have resulted in an increased role of f_0 in realizing the lenis-aspirated contrast. Nevertheless, it is clear from other varieties of Korean as well as data from relatively careful speech that no two of the laryngeal categories have fully merged in VOT, such that there is still something resembling a tripartite VOT contrast in Korean. For example, VOTs of the female model talker in the Korean textbook used by participants in the current study (see Section 2.2) were distinct among the three stop series, as shown in Table 1 (see also the data on native Korean teachers in Chang 2010, 2012).

Table 1. VOT (in ms) of Korean stops as produced by the female model talker in participants' Korean textbook (Seoul National University Language Education Institute 2010). All measures represent VOT of stops in utterance-initial position before the low vowel /a/.

Place of Articulation	Fortis	Lenis	Aspirated
bilabial	10	27	92
denti-alveolar	11	34	94
velar	18	45	122

Thus, regardless of the degree to which lenis and aspirated stops overlap in VOT for individual speakers, it is reasonable to posit that the role of VOT in distinguishing three laryngeal categories in Korean may place a higher burden on this cue in Korean than in languages with a two-way laryngeal contrast (e.g., voiced-voiceless), by far the most common type of spoken language (Maddieson 1984). L2 learners of Korean, likely to have been exposed only to a two-way VOT opposition in their L1, may therefore experience difficulty in acquiring Korean stop laryngeal contrasts, making this set of contrasts useful to examine in a study of L2 perceptual learning.

1.3.2. Vowel Contrasts in Korean

The vowel inventory of modern Korean as spoken in South Korea contains, for most speakers alive today, the monophthongs /i ɛ a u ɨ o ʌ/. A former length contrast has largely disappeared, leaving only short vowels in contemporary varieties (Kim-Renaud 2012). Older descriptions of Korean (e.g., Lee 1993; Sohn 1999; Yang 1992) also include the vowels /e y ø/; however, these vowels are no longer present in the inventory as distinct monophthongs for most speakers. The front rounded vowels have developed into diphthongs (i.e., /ɥi/, /wɛ /; see, e.g., Kim-Renaud 2012), while the former contrast between /e/ and /ɛ / has merged to one mid front vowel phoneme (Ko 2009; Eychenne and Jang 2015, 2018). Thus, apart from several diphthongs (/ɥi wɛ wa wʌ jɛ ja ju jo jʌ ɨi/), modern Korean has a vowel space consisting of seven basic vowel qualities (Chang 2012; Yoon and Kang 2014).

Of interest in the current study are two vowel contrasts that depart from the typologically most common five-vowel inventory of /i e a o u/ (Maddieson 1984)—that between the high back (i.e., non-front) vowels /u/ and /ɨ/ and that between the mid back vowels /o/ and /ʌ/. These contrasts can be described in terms of a difference in phonological labiality or roundedness, as only the former vowel in each pair alternates with labial stops in certain "p-irregular" verbal paradigms and is prohibited from occurring with the rounded on-glide /w/ in diphthongs (Kim-Renaud 1974, 2012; Sohn 1999). Phonetically, however, the two contrasts are realized differently. Whereas /u/ and /ɨ/ differ primarily in terms of the second formant (F_2) and less so in terms of the first (F_1) or third formant (F_3), /o/ and /ʌ/ show substantial differences in both F_1 and F_2 (Table 2; cf. similar data in Chang 2012; Yang 1992; Yoon and Kang 2014; Yoon and Kim 2015).

In short, the two Korean contrasts /u/-/ɨ/ and /o/-/ʌ/, as relatively marked vowel contrasts, may pose a challenge for L2 learners, who may or may not have been exposed to similar contrasts in their L1. Both are therefore examined as target contrasts in the current study.

Table 2. The first three formants (F_1, F_2, F_3) in the target Korean vowels as produced by the female model talker in participants' Korean textbook (Seoul National University Language Education Institute 2010). All measures are in Hz for vowels produced in isolation.

Vowel	F_1	F_2	F_3
/u/	367	800	3149
/ɨ/	376	1628	3056
/o/	390	821	3302
/ʌ/	761	1113	3545

1.4. The Present Study

Given the scarcity of L2 speech research examining the roles of the L1 and IDs in tandem, we conducted a study of L2 perceptual learning of Korean with two goals: (1) examining how L1 transfer and IDs in phonetic sensitivity interact during L2 development, and (2) contributing empirical data on L2 acquisition of Korean, still an underinvestigated L2 despite its increasing popularity as a foreign language worldwide (Byon 2008; Byon and Pyun 2012; Gordon 2015).

With regard to the interaction of transfer with IDs, we hypothesized that L1 transfer in L2 perception would, over time, show an inverse U-shaped pattern instead of a more simple decline. Because the crucial difference between this hypothesis and previous models of L2 learning is in its description of the initial portion of L2 development as showing an increase (as opposed to decrease) in L1 transfer, we call this the TRANSFER RAMP-UP HYPOTHESIS. Note that this hypothesis does not differ from previous models (e.g., OPM; see Section 1.1) in predicting a decline in L1 transfer. Rather, the claim is that the timing of this decline—and, by implication, of L1 transfer effects to begin with—incorporates a delay. That is, the influence of the L1 takes time to "ramp up", peaking at an intermediate, if still early, point in L2 development (see Figure 1).[1] Weak L1 influence at L2 onset thus allows certain non-L1 factors, such as IDs in phonetic sensitivity, to play a greater role earlier (i.e., in the PRE-RAMP-UP stage) than later in L2 development.

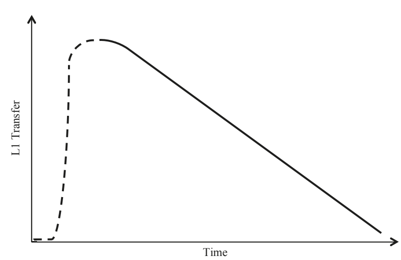


Figure 1. Schematic of the Transfer Ramp-up Hypothesis, showing L1 transfer over time in L2 development. The dotted part of the curve indicates the pre-ramp-up stage and ramp-up to maximal L1 transfer; the solid part, the post-ramp-up stage of declining L1 transfer.

Since the Transfer Ramp-up Hypothesis contradicts several models of L2 learning in positing that L1 transfer does not peak at L2 onset, it is worth explaining the motivation for this hypothesis in more

---

[1] To be more specific, the timescale for the ramp-up is hypothesized to correspond to how long it takes to develop phonological knowledge of the target L2, which is likely to differ depending on the L2 as well as the L1 background of the learner (i.e., the particular pattern of phonological alignment between the L1 and L2). However, in general, we believe that the timescale of the ramp-up to high L1 transfer will be short—less than the five-week interval of L2 learning we observed (see Section 2.2)—because, especially in an instructed L2 context, the phonological system of the L2 (at least, the basic phonological inventory) is one of the first aspects of the language to be acquired, supported by learning the orthographic system in the case of alphabetic orthographies. This is why in Figure 1 the ramp-up is represented as a steep, as opposed to a shallow, incline.

detail. In short, this view follows from two principles alluded to above. The first is that language users maintain access to the perceptual resources that supported acquisition of their L1; the second is that the nature of L2 perception changes with L2 experience. Together, these principles set the stage for L2 perception ab initio, at which point nothing is yet known about the L2, to involve postponing L1 transfer in favor of a language-universal mode of perception. However, the development of a phonological framework for the L2 during L2 learning eventually provides a linguistic basis for mapping L2 speech to the L1, thus encouraging transfer at later stages of learning.

Following from the Transfer Ramp-up Hypothesis, two predictions were tested in the present study. In regard to ab initio perception of the L2, it was predicted that there would be no significant effect of learners' L1 background in a perceptual task not requiring knowledge of L2 phonological categories (P1). That is, performance in such a task was expected to reflect primarily IDs in phonetic sensitivity; therefore, taking ID dimensions as intrinsic properties of the learner, we expected to find a wide range in performance for learners of all L1 backgrounds, and little to no predictive value of linguistically relevant dimensions of a learner's L1. On the other hand, in regard to L2 perception after L2 learning, it was predicted that there would be a significant effect of L1 background, at least in a perceptual task drawing upon knowledge of L2 categories (P2).

Although these predictions were general, and thus applied both to stop and to vowel perception, the linguistically relevant dimensions of a learner's L1 were different for stops and vowels. In the case of stops, the centrality of VOT in distinguishing Korean stop laryngeal categories led us to focus on the number and type of VOT oppositions as the linguistically relevant dimension of the L1. In particular, given that Korean laryngeal categories are all characterized by positive VOTs, we predicted that L1s with VOT oppositions on the positive side of the VOT space (e.g., short- vs. long-lag) would lead to the most beneficial transfer. In the case of vowels, we focused on the occurrence of the target L2 contrasts as the linguistically relevant dimension of the L1 (as a proxy for helpful acoustic oppositions in the back portion of the vowel space), predicting that L1s with one or both of the target contrasts would result in more beneficial transfer than L1s containing neither contrast.

Because this is one of the first studies to examine L2 perceptual learning of Korean by learners from multiple L1 backgrounds, we endeavored to broaden the empirical contribution, as well as the generalizability of the results, by adopting an inclusive approach to the research. As such, we admitted into the study all learners who were eligible rather than targeting a few specific L1 backgrounds. The resulting diversity of L1 backgrounds, some of which are represented by only one participant, naturally presents a challenge for examining the effects of L1 background; we address this by using a group-based analysis, as described in the next section.

2. Methods

The study received ethical approval from the Institutional Review Board (Pukyong National University) under approval code 1041386-202008-HR-47-02. To determine the number of participants to recruit for the study, we carried out a power analysis anticipating multiple regression models with up to eight coefficients apart from the intercept (accounting for group and category predictors consisting of up to three levels, along with interaction coefficients) and assuming 80% power, an alpha level of 0.05, and a model r^2 of 0.24 (based on exploratory modeling of pilot data). Using pwr.f2.test() in the pwr package (Champely 2018) in R (R Development Core Team 2020), we determined the target number of participants to be approximately 56, so we recruited participants until we reached a final sample of at least 56 participants.

2.1. Participants and Groups

A total of 59 adult L2 learners of Korean participated in the study, with two excluded from analysis due to failure to complete the study or outlier performance (i.e., lower than 10% accuracy) in the posttest portion (see Section 2.4.2). The 57 participants in the final analysis (43 female; M_{age} = 21.5 yr, *SD* 4.5) came from 10 L1 backgrounds: North American and British English (*n* = 22), Mandarin Chinese

(n = 14), Finnish (n = 5), Swedish (n = 5), Slovenian (n = 3), Castilian Spanish (n = 2), European French (n = 2), European Portuguese (n = 2), Malay (n = 1), and Turkish (n = 1). They were sorted into groups according to crucial features of their L1 hypothesized to influence perception of the target L2 contrasts—namely, type of VOT contrast and type of vowel inventory.

For the stop study, learners were assigned to one of three groups based on the VOT opposition in their L1 stops: LEAD-SHORT (e.g., [b]-[p]), SHORT-LONG (e.g., [p]-[p^h]), or LEAD-LONG (e.g., [b]-[p^h]). The lead-short group included the L1 speakers of Finnish, French, Malay, Portuguese, Slovenian, and Spanish; the short-long group, English and Mandarin; and the lead-long group, Swedish and Turkish. Using the VOT ranges in Keating (1984), the group classifications were made on the basis of published descriptions of the respective L1s, which suggest that the L1s in the lead-short, short-long, and lead-long groups contrast, respectively, lead (i.e., negative) vs. short-lag VOT (Cruz-Ferreira 1995; Fougeron and Smith 1993; Herrity 2000; Martínez-Celdrán et al. 2003; Shahidi and Aman 2011; Suomi et al. 2008), short- vs. long-lag VOT (Duanmu 2007; Labov et al. 2006; Ladefoged 1999; Roach 2004), and lead vs. long-lag VOT (Helgason and Ringen 2008; Öğüt et al. 2006). Background data on the groups are summarized in Table 3. The groups did not differ significantly in age [$|t|$s < 1.7, ps > 0.05].

Table 3. Background information on participant groups in the stop study.

Group (L1 VOT Type)	n	M_{age} (yr)	L1s Represented
lead-short	15 (9f, 6m)	23.5	Finnish, French, Malay, Portuguese, Slovenian, Spanish
short-long	36 (29f, 7m)	20.6	English, Mandarin
lead-long	6 (4f, 2m)	21.8	Swedish, Turkish

For the vowel study, learners were assigned to one of two groups based on whether their L1 contained none or some of the target vowel contrasts—namely, /u/-/ɨ/ and /o/-/ʌ/. The NO-CONTRAST group included the L1 speakers of Finnish, Mandarin, Spanish, and Swedish; the SOME-CONTRAST group, English, French, Malay, Portuguese, Slovenian, and Turkish. As in the stop study, these group classifications were made on the basis of published descriptions of the respective L1s, which suggest that the L1s in the no-contrast group contain neither contrast (Bradlow 1995; Eklund and Traunmüller 1997; Iivonen and Harnud 2005; Lee and Zee 2003) whereas the L1s in the some-contrast group contain oppositions resembling one or both of the contrasts (Clynes and Deterding 2011; Escudero et al. 2009; Hillenbrand et al. 1995; Jurgec 2005; Kiliç et al. 2004; Strange et al. 2007). Because we relied on published descriptions and there is variation in the transcription conventions used for different languages, we took a somewhat liberal approach to looking for a given contrast; for example, for /u/-/ɨ/, we looked not only for /ɨ/ but also for /ɯ/. Table 4 summarizes the background data on the two groups in the vowel study, which did not differ significantly in age [$t(55) = 1.239$, $p > 0.05$].

Table 4. Background information on participant groups in the vowel study.

Group (L1 Inventory Type)	n	M_{age} (yr)	L1s Represented
no contrast	26 (21f, 5m)	22.3	Finnish, Mandarin, Spanish, Swedish
some contrast	31 (21f, 10m)	20.8	English, French, Malay, Portuguese, Slovenian, Turkish

Participants in all L1 groups tended to have considerable experience with additional languages, such that the majority had been exposed to a type of VOT contrast and/or vowel inventory different from that in their L1. All participants who were not L1 English speakers were also proficient in English (to a level sufficient for college coursework), and most (82%) reported knowledge of one or more other languages (e.g., Cantonese, German, Italian, Japanese, Polish, Russian, Serbo-Croatian, Tagalog, Thai). Given this self-reported multilingualism, we examined whether there were disparities in "diversifying" types of additional language (Ln) exposure across groups—which could potentially result in one group having an inherent advantage in either study—by coding participants in terms of the number of their L*n*s (including English) that would have provided exposure to a type of VOT contrast (or vowel inventory) different from that in their L1. Between-group comparisons on this dimension revealed no significant difference between the no-contrast and some-contrast groups in the vowel study [$t(55) = 1.210$, $p > 0.05$]; however, in the stop study, the short-long group reported significantly fewer “diversifying” L*n*s than the other two groups [$|t|$s > 3.1, *p*s < 0.01], while the lead-short and lead-long groups did not differ from each other in this respect [$t(19) = 0.156$, $p > 0.05$]. The disadvantage of the short-long group in terms of L*n* exposure—largely due to the fact that, unlike the lead-short and lead-long groups, participants in this group did not have English to count as a “diversifying” L*n*—was not, in fact, reflected in systematically lower L2 Korean performance (cf. Section 4); therefore, we assume that L*n* exposure, at least operationalized in terms of number of L*n*s, did not have a detectable impact on the relative patterning of groups on L2 Korean.

2.2. L2 Learning Context

Participants were recruited in 2013–2014 from an international summer program at a university in Seoul, where they were enrolled in an elementary Korean course as well as other courses. The Korean course lasted five weeks and consisted of three hours of instruction per day for 3–4 days a week (18 total class meetings), amounting to a total of 54 contact hours. Consistent (≥90%) attendance was required to pass the course, and every participant received at least a passing grade, meaning that all received a similarly large amount of classroom instruction in the L2. Participants’ non-language courses were conducted in English. Furthermore, students in the program resided in a campus dormitory and, besides contact with Korean students in extracurricular activities, communicated with other students primarily in English. Therefore, in spite of their residence in an L2 environment, participants were not generally immersed in the L2 outside of the classroom.

Although the participants came from different Korean classes split among a team of five instructors (average class size < 15), each class used the same syllabus and teaching materials and followed the same instructional format with English as the main language of instruction. Additionally, the allotment of classroom time to content was uniform across classes, with the first 10 hours allotted to orthographic and phonemic familiarization (i.e., reading and writing of the Korean alphabet). Thus, by the end of the first week of classes, students had mostly finished the portion of the course focused on spelling and pronunciation. Nevertheless, the posttest experiment was not conducted until five weeks later (i.e., at the end of the course) so that participants would be maximally comfortable with using Korean graphemes to provide their responses in this experiment.

2.3. Materials

Stimuli in both the pretest and posttest experiments were designed to test perception of the same Korean phonological contrasts tested in Jung and Kwon (2010): the three-way stop laryngeal contrast and two vowel contrasts, /u/-/ɨ/ and /o/-/ʌ/. Thus, the stimuli included all nine stops (lenis /p t k/, fortis /p* t* k*/, aspirated /p^h t^h k^h/), which were put before /a/, and four lone vowels (/u ɨ o ʌ/).

The stimuli were prepared as follows. First, they were produced by a female native Korean speaker in a standard formal register within the carrier sentence /__-ka is*ɨpnita/ “There is (a) __”. For example, for /k/, the target syllable was /ka/, and this was uttered in the sentence /ka-ka is*ɨpnita/. The set of 13 sentences (9 target stops + 4 target vowels) was arranged in a random order and recorded three times. The recordings were made at 22.1 kHz and 16 bps, in a soundproof booth using a CSL 4500 recording device and a Shure SM48 dynamic mic.

After the sentences were recorded, they were edited in Praat (Boersma and Weenink 2013) to isolate the target syllables. Each syllable was cut out from its carrier sentence and saved in a separate file. All tokens were then inspected auditorily, and one token of each syllable was chosen for the perception stimuli (usually the second token, unless this token sounded unnatural). The resulting 13 stimuli were then randomized and submitted for perceptual evaluation to five native Korean-speaking judges to confirm their quality. The judges were asked to listen to and identify the stimuli (by transcribing them in Korean orthography), as well as rate their confidence in each identification judgment (on a 1–5 scale, 1 being "least confident" and 5 being "very confident"). The rate of correct identification was 100% and the mean confidence rating was 4.9, suggesting that the stimuli were highly intelligible, as well as comprehensible, and thus suitable for use in this study.

2.4. Procedure

Both the pretest and posttest experiments were carried out in a quiet classroom equipped similarly to the rooms for participants' Korean classes. Participants were tested in a group, using the classroom's audio speakers (which were mounted all around the room) and paper answer sheets. Test materials for the pretest and posttest are publicly accessible via the Open Science Framework (OSF) at https://osf.io/rkxdh/.

2.4.1. Pretest: Oddball Discrimination

The pretest experiment was conducted before the beginning of participants' Korean course and was meant to investigate the degree to which these ab initio learners could already perceive the target sounds. Because at this point in time participants had no knowledge of Korean, an identification task was eschewed in favor of a relatively difficult discrimination task (namely, oddball discrimination) with reduced stimulus variability (i.e., only one token of each target syllable) to prevent the task from being overly difficult. Prior to beginning the experiment, participants were told that the sounds they were going to hear were Korean speech sounds (so as to preclude a nonlinguistic mode of auditory perception), and they then completed a short practice session consisting of three trials with stimuli different from the test stimuli to familiarize them with the task.

Each trial of the oddball discrimination task presented a three-item sequence of auditory stimuli to the participant, who had to indicate which (if any) of the three items was different from the other two (i.e., the oddball). For example, one trial presented the sequence /ka/-/k*a/-/k*a/, and the correct answer on this trial was /ka/ (= item #1). The items in a test sequence were separated by an inter-item interval of 700 ms, and each sequence was played twice before participants had to respond. Participants responded by circling the number on their answer sheet corresponding to the serial position of the oddball or, alternatively, one of two other options: "all same" (if no item was different from the other two) and "all different" (if every item was unique). The inter-trial interval was 5 s. Trials were ordered randomly but blocked by contrast type, such that all trials testing laryngeal contrasts were presented before trials testing vowel contrasts. For each of the nine laryngeal contrasts and two vowel contrasts, two test sequences were included (e.g., /p*a/-/p^{h}a/-/p*a/ and /p^{h}a/-/p^{h}a/-/p*a/ for the /p*/-/p^{h}/ contrast), with the position of the oddball distributed across the three possibilities in a ratio of 1:2:2.5. Each test sequence was iterated twice, for a total of 44 trials (11 target contrasts × 2 test sequences × 2 repetitions).

To check that the pretest was in fact reliable, we calculated split-half reliability by randomly dividing the dataset from the pretest into two subsets according to contrast type (i.e., each subset containing half of the items, randomly selected, for each of the five contrast types: lenis vs. fortis, lenis vs. aspirated, fortis vs. aspirated, /o/ vs. /ʌ/, /u/ vs. /ɨ/). This calculation suggested that the pretest had good reliability (Cronbach's $\alpha = 0.82$).

2.4.2. Posttest: Forced-Choice Identification

The posttest experiment was conducted at the end of the Korean course and was meant to examine participants' (relative) degree of success in perceptually acquiring the target Korean sounds. By this point, participants had spent a considerable amount of time learning Korean, including the alphabet, so it was assumed they were familiar enough with its phonological categories and alphabet to perform an orthographic labeling task. Therefore, in contrast to the pretest, the posttest experiment used the identification paradigm to provide a measure more closely reflecting the task of real-world speech perception. Consequently, it should be noted that absolute accuracy levels in the pretest and posttest cannot be directly compared to each other; however, such a comparison is not needed to address our research questions, which concern relative levels of performance across L1 backgrounds rather than absolute accuracy. Prior to beginning this experiment, participants were once again told that they were going to hear Korean speech sounds, and then they completed a short practice session consisting of three trials using stimuli different from the test stimuli to familiarize them with the task.

The identification task was a ten-alternative forced-choice (10AFC) task for trials testing stops and a six-alternative forced-choice (6AFC) task for trials testing vowels. The response options for stop trials were /pa p*a pʰa ta t*a tʰa ka k*a kʰa/ and the distractor option "other". The response options for vowel trials were /u ɨ o ʌ/ and the distractor options /a e/. On each trial, one auditory stimulus consisting of a target syllable in isolation was played twice before participants had to respond. Participants responded by circling the option on their answer sheet (written in Korean orthography) that matched what they had just heard. The inter-trial interval was 5 s. As in the pretest, trials were ordered randomly but blocked by contrast type, such that all trials testing stops were presented before the trials testing vowels. For each of the 13 target syllables, there were three repetitions (of the same token), for a total of 39 trials.

To check that the posttest was reliable, we again calculated split-half reliability by randomly dividing the dataset from the posttest into two subsets according to category type (i.e., each subset containing half of the items, randomly selected, for each of five category types: lenis stops, fortis stops, aspirated stops, mid vowels, high vowels). This calculation suggested that, like the pretest, the posttest also had good reliability (Cronbach's $\alpha = 0.82$).

2.5. Statistical Analysis

In both the stop study and the vowel study, the likelihood of an accurate response in the pretest and posttest experiments was analyzed using mixed-effects logistic regression. To maximize the stability and generalizability of the final models, a two-stage modeling process was used, consisting of initial exploration of potential predictors (in models with lone fixed effects) followed by incremental model building, with model comparisons conducted via likelihood-ratio tests. This process resulted in four final models, one for the pretest and posttest in each of the two studies. All data from the pretest and posttest are publicly accessible via the OSF at https://osf.io/wgcen/.

In the first stage of modeling, we examined a series of models built with single fixed-effect predictors to get a sense of the informativeness of each fixed effect on its own. The random-effects structure in all models comprised intercepts by Participant and Stimulus, except in case there was only one stimulus per L2 category (i.e., in the model of posttest vowel identification).[2] The fixed effects explored in pretest models were the participant's L1 group (Group; treatment-coded with reference level "lead-short" in the stop study and "no contrast" in the vowel study) and the contrast tested on that trial (Contrast; treatment-coded with reference level "lenis-fortis" in the stop study and "/u/-/ɨ/" in the vowel study). The fixed effects explored in posttest models were Group (coded as above), the category tested on that trial (Category; treatment-coded with reference level "lenis" in the stop study and "/u/" in the vowel study), and the participant's overall accuracy in the pretest (PretestAcc; centered and standardized).

In the second stage of modeling, we built a full model on each dataset (e.g., pretest trials in the stop study) with a view toward providing a strong test of the presence of a Group effect. Starting from a base model containing the random-effects structure from above, all potential fixed predictors besides Group were examined first (in decreasing order of informativeness, as indicated by AIC values in the single-predictor models from the first stage of modeling), including all possible interaction terms.

2 A more complex random-effects structure including random slopes was not used in either stage of modelling because models with such a random-effects structure failed to converge or, alternatively, showed signs of overparameterization and/or less stable fit.

Once these predictors were tested and all those which failed to significantly improve the model were removed, the Group term was added to see if it significantly improved the model. Thus, for each dataset, the model improvement (or lack thereof) resulting from the addition of the Group term is interpreted as evidence of the presence/absence of an effect of L1 background.

3. Results

3.1. Individual Differences at Pretest

To check whether IDs were confounded with L1 background (leading to unbalanced levels of IDs across the various L1-based groups), we first examined the range and distribution of variation in pretest performance (i.e., global discrimination accuracy) within each L1 group. Recall that there were three groups in the stop study and two groups in the vowel study; therefore, there were three between-group comparisons to be made in the stop study and one in the vowel study.

As shown in Figure 2, there was some variation across groups in the shape of the distribution in pretest accuracies (e.g., in the stop study, a longer lower tail in the lead-short group vs. a shorter lower tail in the short-long group); however, paired comparisons via Welch-corrected two-sample *t*-tests provided little evidence of a systematic disparity in the degree of IDs between any two groups. For the groups in the stop study, the range in pretest accuracy was 22–97%, 50–100%, and 47–89%, respectively, for the lead-short, short-long, and lead-long groups. None of the between-group differences were significant [$|t|$s < 2.026, ps > 0.05]. As for the groups in the vowel study, the range in pretest accuracy was 47–97% and 22–100%, respectively, for the no-contrast and some-contrast groups. Here, too, the between-group difference was not significant [$t(54.7) = -0.502$, $p > 0.05$]. Given these results, we conclude that IDs in phonetic sensitivity did not simply reflect participants' diverse L1 backgrounds. In fact, there were substantial, and not significantly different, levels of individual variation in all of the L1 groups discussed below.

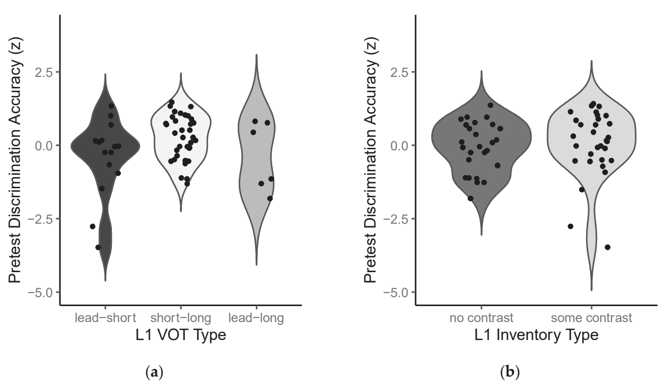


Figure 2. Violin plots showing the probability density of pretest discrimination accuracies (standardized) in each L1 group in (**a**) the stop study and (**b**) the vowel study. Superimposed points (jittered along the *x*-axis) represent individual participants.

3.2. Study 1: Stop Perception

Analysis of pretest performance in the stop study revealed considerable variation in discrimination accuracy across the different L2 laryngeal contrasts, but no systematic effect of the learner's L1 group. As shown in Figure 3, accuracy was well above chance level (=20%) in all cases, but tended to be lower on the lenis-fortis contrast (M_{acc} = 58%) than on the other two contrasts (fortis-aspirated: M_{acc} = 85%; lenis-aspirated: M_{acc} = 81%). This was reflected in the modeling results, which revealed a significant effect of Contrast [$\chi^2(2) = 13.769$, $p < 0.01$] but no effect of Group [$\chi^2(2) = 5.604$, $p > 0.05$]. The final model (shown in Table 5) confirmed that, compared to the lenis-fortis contrast (represented in the intercept), the fortis-aspirated and lenis-aspirated contrasts were both significantly more likely to be discriminated accurately [βs > 1.476, ps < 0.001].

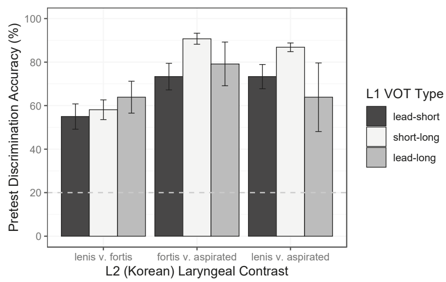


Figure 3. Pretest performance in the stop study (accuracy in oddball discrimination), by L2 laryngeal contrast and L1 group (i.e., VOT type). Error bars indicate *SE* of the mean over participants; the dotted line marks chance-level performance (20%).

Table 5. Fixed-effect coefficients in the final model of pretest stop discrimination [$N = 2052$, log-likelihood $= -958.7$]. Significance code: *** $p < 0.001$.

Predictor	β	*SE*	z	p	
(Intercept)	0.421	0.321	1.309	0.191	
Contrast: fortis-aspirated	1.730	0.416	4.162	<0.001	***
Contrast: lenis-aspirated	1.477	0.414	3.565	<0.001	***

Analysis of posttest performance in the stop study revealed both variation in identification accuracy across the different L2 laryngeal categories as well as an advantage of the short-long group—and, to a lesser extent, the lead-short group—over the lead-long group. As shown in Figure 4, with the exception of the lead-long group on fortis stops, accuracy was well above chance level (i.e., 10%); nevertheless, accuracy tended to be lower on fortis stops (M_{acc} = 51%) than on the lenis (M_{acc} = 68%) or aspirated stops (M_{acc} = 62%), especially for the lead-short and lead-long groups. These patterns were reflected in the modeling results, which showed no effect of PretestAcc [$\chi^2(1) = 0.834$, $p > 0.05$] but a significant effect of Category [$\chi^2(2) = 13.770$, $p < 0.01$], of Group [$\chi^2(2) = 8.564$, $p < 0.05$], and of the Category × Group interaction [$\chi^2(4) = 36.291$, $p < 0.001$].

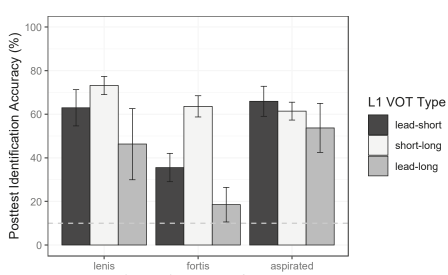


Figure 4. Posttest performance in the stop study (accuracy in 10AFC identification), by L2 laryngeal category and L1 group (i.e., VOT type). Error bars indicate *SE* of the mean over participants; the dotted line marks chance-level performance (10%).

The nature of the interaction between Category and Group can be seen in the final model of posttest accuracy (Table 6). Relative to accuracy on lenis stops, the lead-short group was significantly less likely to be accurate on fortis stops [$\beta = -1.394$, $p < 0.001$] but not aspirated stops [$\beta = 0.163$, $p > 0.05$], and the lead-long group showed a similar pattern, reflected in non-significant interaction coefficients. The short-long group, however, showed a different pattern, with a much smaller decrease in accuracy on fortis stops vis-a-vis lenis stops and a slight decrease in accuracy on aspirated stops (contrasting with the slight increase observed in the other two groups); these differences were reflected in a significant positive interaction coefficient for fortis stops [$\beta = 0.860$, $p < 0.05$] and a significant negative interaction coefficient for aspirated stops [$\beta = -0.810$, $p < 0.05$].

Table 6. Fixed-effect coefficients in the final model of posttest stop identification [$N = 1539$, log likelihood $= -885.7$]. Significance codes: * $p < 0.05$, *** $p < 0.001$.

Predictor	β	*SE*	*z*	*p*	
(Intercept)	0.630	0.346	1.822	0.069	
Category: fortis	−1.394	0.288	−4.847	<0.001	***
Category: aspirated	0.163	0.287	0.568	0.570	
Group: short-long	0.617	0.414	1.492	0.136	
Group: lead-long	−0.792	0.642	−1.234	0.217	
Category: fortis × Group: short-long	0.860	0.340	2.530	0.011	*
Category: aspirated × Group: short-long	−0.810	0.339	−2.388	0.017	*
Category: fortis × Group: lead-long	−0.310	0.583	−0.532	0.594	
Category: aspirated × Group: lead-long	0.223	0.522	0.428	0.669	

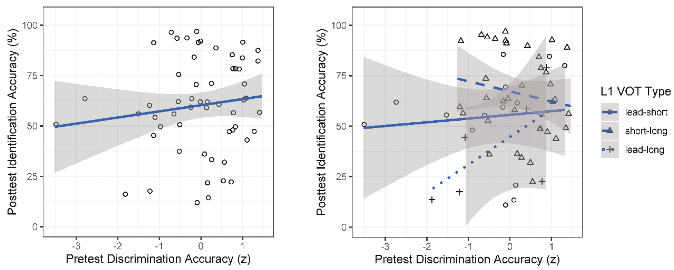
In addition to analyzing accuracy, we also inspected the pattern of errors in the posttest to check whether they involved laryngeal category, place (of articulation), or both. As shown in Figure 5 (see also Figures A1–A3 in Appendix A), errors for every stimulus mostly involved laryngeal category only. The two exceptions were /p^{h}a/ and /ta/, which each elicited a high proportion of errors involving place. Crucially, however, errors involving laryngeal category only were by far the most common error type, both overall (83%) and across groups (77–88%).

	/pa/	/p*a/	/p^{h}a/	/ta/	/t*a/	/t^{h}a/	/ka/	/k*a/	/k^{h}a/	other
/pa/ (53)		**45**	42	2	9	0	2	0	0	0
/p*a/ (87)	**62**		26	3	7	1	0	0	0	0
/p^{h}a/ (57)	26	23		7	5	**32**	0	0	4	4
/ta/ (52)	12	8	6		**35**	29	10	0	0	2
/t*a/ (74)	1	7	1	38		**49**	1	1	0	1
/t^{h}a/ (71)	3	0	1	**61**	30		0	1	3	1
/ka/ (61)	3	3	2	2	0	0		**49**	41	0
/k*a/ (87)	0	3	0	1	0	5	**51**		40	0
/k^{h}a/ (68)	1	0	0	1	3	1	**53**	38		1

Figure 5. Confusion matrix of errors in the posttest for stop items (vertical = stimuli, horizontal = responses). Each cell shows the percentage of all errors on the given stimulus represented by the given response (rows may not add to 100% due to rounding); the most common error type for each stimulus is bolded. The total number of errors for each stimulus (across all groups) is shown in parentheses.

Although our modeling results suggested that pretest performance was not a predictor of posttest stop perception, we further examined the relationship between the two through Pearson's correlations to provide additional evidence of the (non)predictiveness of pretest performance. These analyses revealed no significant correlations between pretest and posttest accuracy on stops, overall [$r(55) = 0.13$, $p > 0.05$] or within any individual group: lead-short [$r(13) = 0.10$, $p > 0.05$], short-long [$r(34) = -0.18$,

$p > 0.05$], or lead-long [$r(4) = 0.64$, $p > 0.05$] (see Figure 6). In short, results of the stop study produced some evidence of L1 transfer, but crucially only following extensive L2 exposure; moreover, there was no evidence of a link between preexisting phonetic sensitivity (as measured in the pretest) and L2 stop perception following L2 learning (as measured in the posttest).



(**a**) (**b**)

Figure 6. Individual posttest accuracy in the stop study by pretest performance, with regression lines (**a**) over all L1 groups together, and (**b**) over each L1 group separately. Points are jittered; the shaded areas indicate the 95% confidence interval around the regression line.

3.3. Study 2: Vowel Perception

Analysis of pretest performance in the vowel study revealed ceiling-level discrimination accuracy for both L2 vowel contrasts and across both L1 groups. As shown in Figure 7, accuracy was well above chance in all cases and did not differ between the /u/-/ɨ/ and /o/-/ʌ/ contrasts (M_{acc} = 93% for both). Modeling results showed no effect of Contrast [$\chi^2(1) = 0$, $p > 0.05$] or of Group [$\chi^2(1) = 0.040$, $p > 0.05$], so the final model [$N = 456$, log-likelihood $= -83.5$] contained no fixed predictors. Consistent with the high accuracies in Figure 6, the intercept in this model indicated that L2 vowel contrasts were, overall, discriminated with significantly better than 50-50 odds [$\beta = 6.983$, $z = 4.456$, $p < 0.001$].

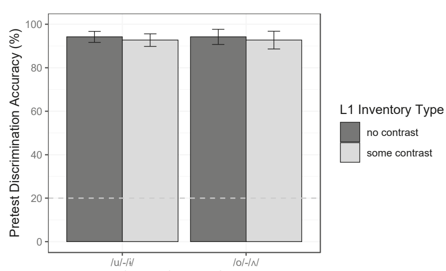


Figure 7. Pretest performance in the vowel study (accuracy in oddball discrimination), by L2 vowel contrast and L1 group (i.e., vowel inventory type). Error bars indicate *SE* of the mean over participants; the dotted line marks chance-level performance (20%).

Compared to pretest performance, there was more variation in posttest performance, but accuracies were generally high, especially relative to stop identification rates. There were some differences among vowels, with /ɨ/ showing the highest accuracies and /ʌ/ the lowest. Crucially, there was also a group difference: the some-contrast group outperformed the no-contrast group, both overall (no contrast: M_{acc} = 74%; some-contrast: M_{acc} = 88%) and on every vowel category, as shown in Figure 8. Thus, modeling results showed a significant effect of Category [$\chi^2(3) = 52.729$, $p < 0.001$] and of Group [$\chi^2(1) = 4.233$, $p < 0.05$] though no effect of the Category × Group interaction [$\chi^2(3) = 0.955$, $p > 0.05$]. Additionally, as in the stop study, there was no effect of PretestAcc [$\chi^2(1) = 0.395$, $p > 0.05$]. The final model (Table 7) indicated that, compared to /u/, /ɨ/ was significantly more likely to be identified accurately [$\beta = 2.090$, $p < 0.001$] while /Λ/ was significantly less likely to be identified accurately [$\beta = -0.768$, $p < 0.05$]; furthermore, the some-contrast group was significantly more likely to be accurate than the no-contrast group [$\beta = 1.210$, $p < 0.05$].

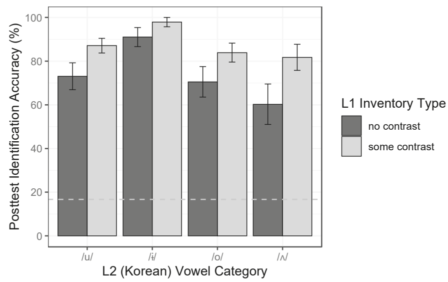


Figure 8. Posttest performance in the vowel study (accuracy in 6AFC identification), by L2 vowel category and L1 group (i.e., vowel inventory type). Error bars indicate *SE* of the mean over participants; the dotted line marks chance-level performance (17%).

Table 7. Fixed-effect coefficients in the final model of posttest vowel identification [$N = 684$, log-likelihood = −248.0]. Significance codes: * $p < 0.05$, *** $p < 0.001$.

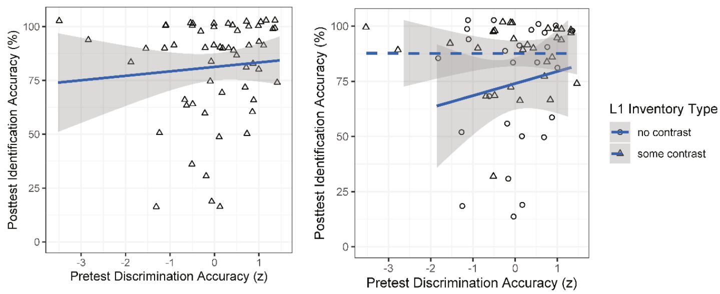
Predictor	β	*SE*	*z*	*p*	
(Intercept)	1.611	0.474	3.395	<0.001	***
Category: /ɨ/	2.090	0.482	4.336	<0.001	***
Category: /o/	−0.276	0.333	−0.829	0.407	
Category: /Λ/	−0.768	0.326	−2.356	0.018	*
Group: some-contrast	1.210	0.579	2.091	0.037	*

As in the study of stop perception, here too we inspected the pattern of errors in the posttest to see which vowels were most confusable with each other. As shown in Figure 9 (see also Figures A4 and A5 in Appendix A), /u/ was most often confused with /o/, /ɨ/ with /u/ (although there were very few errors on /ɨ/ overall), /o/ with /Λ/, and /Λ/ with /a/. Relatively few errors involved confusion with the vowel /i/ or /e/. In short, vowel identification errors tended to involve confusion with a vowel sharing a specification for roundedness/labiality and/or height, which was unsurprising; however, predominant vowel confusions were not symmetrical (e.g., /o/ was most often misidentified as /Λ/ but /Λ/ was not most often misidentified as /o/).

	/u/	/ɨ/	/o/	/ʌ/	/a/	/e/
/u/ (33)		12	**61**	27	0	0
/ɨ/ (9)	**56**		33	0	0	11
/o/ (38)	16	8		**74**	0	3
/ʌ/ (48)	8	2	27		**63**	0

Figure 9. Confusion matrix of errors in the posttest for vowel items (vertical = stimuli, horizontal = responses). Each cell shows the percent of all errors on the given stimulus represented by the given response (rows may not add to 100% due to rounding); the most common error type for each stimulus is bolded. The total number of errors for each stimulus (across all groups) is shown in parentheses.

As in the stop study, we further examined the relationship between pretest performance and posttest vowel perception via Pearson's correlations. In line with the modeling results as well as the lack of correlations seen in the stop study, these analyses revealed no significant correlations between pretest performance and posttest accuracy on vowels, overall [$r(55) = 0.09$, $p > 0.05$] or within either individual group: no-contrast [$r(24) = 0.16$, $p > 0.05$] or some-contrast [$r(29) = -0.002$, $p > 0.05$] (see Figure 10). In short, results of the vowel study were consistent with those of the stop study: again, we found evidence of L1 transfer only in the posttest and no evidence of a link between preexisting phonetic sensitivity and L2 vowel perception after L2 learning.



(**a**) (**b**)

Figure 10. Individual posttest accuracy in the vowel study by pretest performance, with regression lines (**a**) over all L1 groups together, and (**b**) over each L1 group separately. Points are jittered; the shaded areas indicate the 95% confidence interval around the regression line.

4. Discussion

In summary, evidence of L1 transfer was not found in L2 perception of Korean ab initio, but was found after a prolonged period of L2 learning. Results of the pretest showed, as expected, substantial individual differences (IDs) in learners' preexisting phonetic sensitivity to the target stop and vowel contrasts, but no systematic effect of the crucial L1 features hypothesized to affect L2 perception of these contrasts. In contrast, results of the posttest showed a significant effect of these L1 features, both for stops and for vowels, but no effect of phonetic sensitivity as reflected in pretest performance. These findings thus provide support for the first part of the Transfer Ramp-up Hypothesis: rather than peaking at L2 onset, L1 influence takes time to set in during L2 perceptual development (at which point the relative role of IDs in phonetic sensitivity may become smaller).

Although these findings converge with the results of previous studies on L2 Korean (Jung and Kwon 2010; Kwon 2013), they also diverge from the literature supporting the Perceptual Assimilation Model (PAM; Best 1994, 1995), much of which also examined naive listeners of an L2 yet found patterns of ostensible L1 influence in discrimination of L2 contrasts. In our view, the disparity between the current results and PAM-framed results is likely due to different formulations of the dependent measure. That is, whereas studies testing PAM have generally examined L2/nonnative perception by individual contrast (because PAM makes different predictions for different types of L2 contrasts), the present study, which was not focused on testing PAM, evaluated L2 perception mainly in terms of an overall likelihood of accuracy. In fact, when the (pretest) discrimination results are separated out by contrast, there is some indication of a possible L1 effect for certain contrasts (e.g., the short-long group showing the greatest advantage over the lead-short group on the Korean fortis vs. aspirated contrast; see Figure 3), even if not for others (e.g., the no-contrast group showing similar, ceiling-level performance as the some-contrast group on both Korean vowel contrasts; see Figure 7).

Thus, while we interpret the current findings as support for the Transfer Ramp-up Hypothesis, we are also careful to point out that our observation of a low degree of L1 transfer at L2 onset (vis-a-vis later points in L2 development) does not mean that there is no L1 influence at L2 onset. It is not possible to draw such a conclusion on the basis of the evidence in this study, and the Transfer Ramp-up Hypothesis, moreover, does not make this extreme claim. The core claim is rather that there is an increase in L1 transfer during the early part of L2 development. Because this does not entail the total absence of L1 effects at L2 onset, it stands to reason that the L1 may indeed have a detectable effect on the ab initio perception of specific L2 contrasts (e.g., "Single Category" contrasts that assimilate to the same L1 sound; Best 1994), which may amount to an L1 effect that is weak or undetectable overall when multiple L2 contrasts are considered together. The locus of the difference between the present study and previous theoretical frameworks, therefore, is in the observed trajectory of L1 transfer. In particular, our findings contradict the view—made explicit in the Ontogeny Phylogeny Model (OPM) and left implicit in other frameworks such as PAM(-L2)—that L1 influence is at its height at the start of L2 development and only decreases from that point.

Apart from the implications for theories of L1 transfer, the current findings also have implications for views of IDs in L2 acquisition. In particular, they make two contributions to our understanding of IDs in L2 perceptual development. First, given that no effect of L1 background was found in the pretest despite ample variation among learners, the pretest results provide evidence that the magnitude of IDs in phonetic sensitivity to L2 contrasts—at least, to the target L2 Korean contrasts examined in this study—can be relatively large compared to the magnitude of L1 effects. These results thus argue in favor of making the analysis of IDs central to studies of L2 perception (as is becoming increasingly common in the L2 speech literature; see Section 2.2) because IDs may actually prove to be a more powerful predictor of L2 perceptual behavior (at certain points in L2 development) than the more extensively examined factor of L1 background. Second, together with the pretest results, the posttest results provide evidence that the effect of IDs changes over the course of L2 development. Crucially, IDs in phonetic sensitivity were found not to predict L2 perceptual accuracy after L2 learning, which supports a view of L2 perception, at any given point during L2 development, as the outcome of a dynamic interaction between L1 transfer and IDs.

In light of these theoretical implications, it would be remiss of us not to mention the limitations of the present study, which motivate a number of different directions for future research. First, as alluded to in Section 2.4, the composition of the participant sample, influenced in large part by worldwide trends in who elects to study Korean as an L2, was unevenly distributed in terms of L1 backgrounds, necessitating an approach that grouped L1s together rather than analyzing them separately. Although adequate for addressing research questions related to broadly formulated phonological features of an L1, this approach does not lend itself to a nuanced examination of features that may be more language-specific (e.g., palatalization in Russian; pharyngealization in Arabic). It would, therefore, provide additional insight to replicate this type of study with a larger sample in which all L1 backgrounds are robustly represented, allowing for analyses that focus on specific L1s.

Second, as mentioned in Section 2.4.2, the study was designed to use different task paradigms (discrimination and identification) for the pretest and posttest, considering both ecological validity and appropriateness for different stages of L2 learning. Although both paradigms are widely used to measure perceptual ability, and the measures from these paradigms have been shown to be highly correlated with each other at the same point in time (e.g., Pearson's $r > 0.6$ for L2 discrimination and identification of Mandarin tones; Bowles et al. 2016), the fact remains that discrimination performance and identification performance cannot be directly compared to each other, which prevents us from being able to draw conclusions about participants' L2 perceptual development that are truly longitudinal. Thus, it would be useful in future research to collect longitudinal data from the exact same task, which, with some design modifications, could be made appropriate for learners at different stages of L2 development (e.g., use of iconic images for response options in ab initio L2 identification; see Bowles et al. 2016).

Third, this study included observations of two time points in L2 development, whereas at least three are needed to fully demonstrate the inverse U-shaped pattern of L1 transfer postulated in the Transfer Ramp-up Hypothesis (Figure 1). That is, we have only assumed the part of the pattern in which L1 transfer declines at later points in L2 development on the basis of prior findings in the literature (see Section 1.1), since it was not possible to observe a third, later time point in the case of the current participants (who did not necessarily continue learning Korean after the end of their Korean language course). In future work, it would therefore be helpful to track learners further into their trajectory of L2 learning (e.g., in a year-long course of L2 instruction) so as to provide direct evidence of the hypothesized post-ramp-up decline in L1 transfer.

5. Conclusions

In closing, we would like to highlight one of the chief challenges of designing developmental perceptual research such as the present study, and outline a possible approach to addressing this challenge in future research. In our view, truly longitudinal perceptual data (i.e., data from the same individual completing the same perceptual task, including the same auditory stimuli, at different points in time) may not be the ideal data for investigating perceptual change over time, because it is not only development, but also extraneous factors such as familiarity with (or memory of) the test stimuli, that may lead to listeners performing differently in the same task across two points in time. The challenge for future research, therefore, is to identify and control for such extraneous factors appropriately. In the case of test stimuli, for example, one way of addressing the issue of familiarity/memory would be to use similarly constructed, but non-identical, stimulus sets at different time points, normed in advance to be equivalently difficult. Studies designed "pseudo-longitudinally" in this manner would be better able to show change that could be confidently interpreted as reflecting development.

Despite the challenges of incorporating a longitudinal design in developmental perceptual research, however, the need for longitudinal studies in working toward a theory of L2 perceptual learning that incorporates a role for both L1 transfer and individual differences cannot be overemphasized. Given that cross-sectional studies, by their very nature, are ill-equipped to examine the temporal dynamicity of individual difference effects, longitudinal studies, including both the laboratory training approach as well as the classroom learning approach taken in the present study, are uniquely positioned to shed new light on the roles, and interaction, of language-specific and personally-specific variables in L2 perceptual development.

Author Contributions: Conceptualization, C.B.C. and S.K.; Data curation, C.B.C.; Formal analysis, C.B.C.; Funding acquisition, S.K.; Investigation, S.K.; Methodology, C.B.C. and S.K.; Project administration, S.K.; Resources, S.K.; Supervision, S.K.; Validation, C.B.C. and S.K.; Visualization, C.B.C.; Writing—original draft, C.B.C.; Writing—review & editing, C.B.C. and S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Pukyong National University, grant number C-D-2016-0831.

Acknowledgments: The authors gratefully acknowledge Peter Jurgec's expert advice on Slovenian phonetics and phonology and helpful comments and feedback from two anonymous reviewers.

Conflicts of Interest: The authors declare no conflict of interest. The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A

	/pa/	/p*a/	/pʰa/	/ta/	/t*a/	/tʰa/	/ka/	/k*a/	/kʰa/	other
/pa/ (19)		**53**	21	5	21	0	0	0	0	0
/p*a/ (31)	**55**		32	3	10	0	0	0	0	0
/pʰa/ (14)	14	**29**		7	14	**29**	0	0	7	0
/ta/ (14)	7	14	7		**36**	29	0	0	0	7
/t*a/ (25)	4	12	4	28		**44**	4	0	0	4
/tʰa/ (15)	13	0	0	**47**	40		0	0	0	0
/ka/ (17)	6	6	6	0	0	0		**53**	29	0
/k*a/ (31)	0	3	0	3	0	10	**52**		32	0
/kʰa/ (17)	0	0	0	0	12	0	35	**47**		6

Figure A1. Confusion matrix of errors by the lead-short group in the posttest for stop items (vertical = stimuli, horizontal = responses). Each cell shows the percent of all errors on the given stimulus represented by the given response (rows may not add to 100% due to rounding); the most common error type for each stimulus is bolded. The total number of errors for each stimulus (across all participants in the group) is shown in parentheses.

	/pa/	/p*a/	/pʰa/	/ta/	/t*a/	/tʰa/	/ka/	/k*a/	/kʰa/	other
/pa/ (24)		42	**58**	0	0	0	0	0	0	0
/p*a/ (40)	**68**		23	5	5	0	0	0	0	0
/pʰa/ (34)	29	21		6	3	**35**	0	0	0	6
/ta/ (26)	15	0	4		27	**35**	19	0	0	0
/t*a/ (36)	0	0	0	42		**58**	0	0	0	0
/tʰa/ (48)	0	0	2	**63**	29		0	0	4	2
/ka/ (37)	3	0	0	3	0	0		**49**	46	0
/k*a/ (42)	0	0	0	0	0	0	**52**		48	0
/kʰa/ (43)	0	0	0	2	0	0	**65**	33		0

Figure A2. Confusion matrix of errors by the short-long group in the posttest for stop items (vertical = stimuli, horizontal = responses). Each cell shows the percent of all errors on the given stimulus represented by the given response (rows may not add to 100% due to rounding); the most common error type for each stimulus is bolded. The total number of errors for each stimulus (across all participants in the group) is shown in parentheses.

	/pa/	/p*a/	/pʰa/	/ta/	/t*a/	/tʰa/	/ka/	/k*a/	/kʰa/	other
/pa/ (10)		**40**	**40**	0	10	0	10	0	0	0
/p*a/ (16)	**63**		25	0	6	6	0	0	0	0
/pʰa/ (9)	**33**	22		11	0	22	0	0	11	0
/ta/ (12)	8	17	8		**50**	17	0	0	0	0
/t*a/ (13)	0	15	0	**46**		31	0	8	0	0
/tʰa/ (8)	0	0	0	**75**	13		0	13	0	0
/ka/ (7)	0	14	0	0	0	0		**43**	**43**	0
/k*a/ (14)	0	14	0	0	0	7	**43**		36	0
/kʰa/ (8)	13	0	0	0	0	13	25	**50**		0

Figure A3. Confusion matrix of errors by the lead-long group in the posttest for stop items (vertical = stimuli, horizontal = responses). Each cell shows the percent of all errors on the given stimulus represented by the given response (rows may not add to 100% due to rounding); the most common error type for each stimulus is bolded. The total number of errors for each stimulus (across all participants in the group) is shown in parentheses.

	/u/	/ɨ/	/o/	/ʌ/	/a/	/e/
/u/ (21)		14	**48**	38	0	0
/ɨ/ (7)	**57**		43	0	0	0
/o/ (23)	9	13		**78**	0	0
/ʌ/ (31)	10	0	29		**61**	0

Figure A4. Confusion matrix of errors by the no-contrast group in the posttest for vowel items (vertical = stimuli, horizontal = responses). Each cell shows the percent of all errors on the given stimulus represented by the given response (rows may not add to 100% due to rounding); the most common error type for each stimulus is bolded. The total number of errors for each stimulus (across all participants in the group) is shown in parentheses.

	/u/	/ɨ/	/o/	/ʌ/	/a/	/e/
/u/ (12)		8	**83**	8	0	0
/ɨ/ (2)	**50**		0	0	0	**50**
/o/ (15)	27	0		**67**	0	7
/ʌ/ (7)	6	6	24		**65**	0

Figure A5. Confusion matrix of errors by the some-contrast group in the posttest for vowel items (vertical = stimuli, horizontal = responses). Each cell shows the percent of all errors on the given stimulus represented by the given response (rows may not add to 100% due to rounding); the most common error type for each stimulus is bolded. The total number of errors for each stimulus (across all participants in the group) is shown in parentheses.

References

Baker, Wendy, and Pavel Trofimovich. 2006. Perceptual paths to accurate production of L2 vowels: The role of individual differences. *International Review of Applied Linguistics in Language Teaching* 44: 231–50. [CrossRef]

Bergier, Marcin. 2014. The influence of explicit phonetic instruction and production training practice on awareness raising in the realization of stop consonant clusters by advanced Polish learners of English. In *Awareness in Action: The Role of Consciousness in Language Acquisition*. Edited by Andrzej Łyda and Konrad Szcześniak. Heidelberg: Springer, pp. 103–20. [CrossRef]

Best, Catherine T. 1994. The emergence of native-language phonological influences in infants: A perceptual assimilation model. In *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*. Edited by Judith C. Goodman and Howard C. Nusbaum. Cambridge: MIT Press, pp. 167–224.

Best, Catherine T. 1995. A direct realist view of cross-language speech perception. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Edited by Winifred Strange. Baltimore: York Press, pp. 171–204.

Best, Catherine T., and Michael D. Tyler. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*. Edited by Ocke-Schwen Bohn and Murray J. Munro. Amsterdam: John Benjamins Publishing, pp. 13–34. [CrossRef]

Boersma, Paul, and David Weenink. 2013. Praat: Doing Phonetics by Computer. Version 5.3.48. Available online: http://www.praat.org (accessed on 1 May 2013).

Bohn, Ocke-Schwen, and Catherine T. Best. 2012. Native-language phonetic and phonological influences on perception of American English approximants by Danish and German listeners. *Journal of Phonetics* 40: 109–28. [CrossRef]

Bowles, Anita R., Charles B. Chang, and Valerie P. Karuzis. 2016. Pitch ability as an aptitude for tone learning. *Language Learning* 66: 774–808. [CrossRef]

Bradlow, Ann R. 1995. A comparative acoustic study of English and Spanish vowels. *Journal of the Acoustical Society of America* 97: 1916–24. [CrossRef] [PubMed]

Byon, Andrew Sangpil. 2008. Korean as a foreign language in the USA: The instructional settings. *Language, Culture and Curriculum* 21: 244–55. [CrossRef]

Byon, Andrew Sangpil, and Danielle Ooyoung Pyun, eds. 2012. *Teaching and Learning Korean as a Foreign Language: A Collection of Empirical Studies*. Columbus: Ohio State University Press.

Carroll, John B. 1981. Twenty-five years of research on foreign language aptitude. In *Individual Differences and Universals in Language Learning Aptitude*. Edited by Karl C. Diller. Rowley: Newbury House, pp. 83–118.

Champely, Stephane. 2018. Pwr: Basic Functions for Power Analysis [R package]. Version 1.2-2. Available online: http://cran.r-project.org/web/packages/pwr/index.html (accessed on 1 April 2018).

Chandrasekaran, Bharath, Padma D. Sampath, and Patrick C. M. Wong. 2010. Individual variability in cue-weighting and lexical tone learning. *Journal of the Acoustical Society of America* 128: 456–65. [CrossRef] [PubMed]

Chang, Charles B. 2010. The implementation of laryngeal contrast in Korean as a Second Language. *Harvard Studies in Korean Linguistics* 13: 91–104.

Chang, Charles B. 2012. Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics* 40: 249–68. [CrossRef]

Chang, Charles B. 2013. The production and perception of coronal fricatives in Seoul Korean: The case for a fourth laryngeal category. *Korean Linguistics* 15: 7–49. [CrossRef]

Chang, Charles B. 2016. Bilingual perceptual benefits of experience with a heritage language. *Bilingualism: Language and Cognition* 19: 791–809. [CrossRef]

Chang, Charles B. 2018. Perceptual attention as the locus of transfer to nonnative speech perception. *Journal of Phonetics* 68: 85–102. [CrossRef]

Cho, Taehong, Sun-Ah Jun, and Peter Ladefoged. 2002. Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics* 30: 193–228. [CrossRef]

Clynes, Adrian, and David Deterding. 2011. Standard Malay (Brunei). *Journal of the International Phonetic Association* 41: 259–68. [CrossRef]

Cruz-Ferreira, Madalena. 1995. European Portuguese. *Journal of the International Phonetic Association* 25: 90–94. [CrossRef]

Darcy, Isabelle, Hanyong Park, and Chung-Lin Yang. 2015. Individual differences in L2 acquisition of English phonology: The relation between cognitive abilities and phonological processing. *Learning and Individual Differences* 40: 63–72. [CrossRef]

Dewaele, Jean-Marc, and Adrian Furnham. 2000. Personality and speech production: A pilot study of second language learners. *Personality and Individual Differences* 28: 355–65. [CrossRef]

Díaz, Begoña, Cristina Baus, Carles Escera, Albert Costa, and Núria Sebastián-Gallés. 2008. Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language. *Proceedings of the National Academy of Sciences of the United States of America* 105: 16083–88. [CrossRef]

Díaz, Begoña, Holger Mitterer, Mirjam Broersma, and Núria Sebastián-Gallés. 2012. Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access. *Learning and Individual Differences* 22: 680–89. [CrossRef]

Dörnyei, Zoltán, and Peter Skehan. 2003. Individual differences in second language learning. In *The Handbook of Second Language Acquisition*. Edited by Catherine J. Doughty and Michael H. Long. Malden: Blackwell Publishing, pp. 589–630. [CrossRef]

Dörnyei, Zoltán. 2006. Individual differences in second language acquisition. *AILA Review* 19: 42–68. [CrossRef]

Duanmu, San. 2007. *The Phonology of Standard Chinese*, 2nd ed. Oxford: Oxford University Press.

Eckman, Fred R., Abdulla Elreyes, and Gregory K. Iverson. 2001. Allophonic splits in L2 phonology: The question of learnability. *International Journal of English Studies* 1: 21–51.

Eklund, Ingegerd, and Hartmut Traunmüller. 1997. Comparative study of male and female whispered and phonated versions of the long vowels of Swedish. *Phonetica* 54: 1–21. [CrossRef]

Escudero, Paola, Paul Boersma, Andréia Schurt Rauber, and Ricardo A. H. Bion. 2009. A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *Journal of the Acoustical Society of America* 126: 1379–93. [CrossRef] [PubMed]

Eychenne, Julien, and Tae-Yeoub Jang. 2015. On the merger of Korean mid front vowels: Phonetic and phonological evidence. *Malsoriwa Eumseong Gwahak [Phonetics and Speech Sciences]* 7: 119–29. [CrossRef]

Eychenne, Julien, and Tae-Yeoub Jang. 2018. On the role of contrastivity in the development of the /e∼ɛ / merger in Korean. *Journal of East Asian Linguistics* 27: 111–39. [CrossRef]

Flege, James Emil. 1995. Second language speech learning: Theory, findings, and problems. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Edited by Winifred Strange. Baltimore: York Press, pp. 233–72.

Fougeron, Cécile, and Caroline L. Smith. 1993. French. *Journal of the International Phonetic Association* 23: 73–76. [CrossRef]

Golestani, Narly, and Robert J. Zatorre. 2009. Individual differences in the acquisition of second language phonology. *Brain and Language* 109: 55–67. [CrossRef]

Gordon, Larry. 2015. Korean-Language Classes Are Growing in Popularity at U.S. *Colleges*. *Los Angeles Times*. April 1. Available online: https://lat.ms/1GjgEjG (accessed on 1 August 2018).

Hanulíková, Adriana, Dan Dediu, Zhou Fang, Jana Bašnaková, and Falk Huettig. 2012. Individual differences in the acquisition of a complex L2 phonology: A training study. *Language Learning* 62: 79–109. [CrossRef]

Hattori, Kota, and Paul Iverson. 2009. English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *Journal of the Acoustical Society of America* 125: 469–79. [CrossRef]

Helgason, Pétur, and Catherine Ringen. 2008. Voicing and aspiration in Swedish stops. *Journal of Phonetics* 36: 607–28. [CrossRef]

Herrity, Peter. 2000. *Slovene: A Comprehensive Grammar*. London: Routledge.

Hillenbrand, James, Laura A. Getty, Michael J. Clark, and Kimberlee Wheeler. 1995. Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America* 97: 3099–111. [CrossRef]

Holliday, Jeffrey J. 2014. The perceptual assimilation of Korean obstruents by native Mandarin listeners. *Journal of the Acoustical Society of America* 135: 1585–95. [CrossRef]

Holliday, Jeffrey J. 2015. A longitudinal study of the second language acquisition of a three-way stop contrast. *Journal of Phonetics* 50: 1–14. [CrossRef]

Holliday, Jeffrey J. 2016. Second language experience can hinder the discrimination of nonnative phonological contrasts. *Phonetica* 73: 33–51. [CrossRef] [PubMed]

Idemaru, Kaori, Lori L. Holt, and Howard Seltman. 2012. Individual differences in cue weights are stable across time: The case of Japanese stop lengths. *Journal of the Acoustical Society of America* 132: 3950–64. [CrossRef]

Iivonen, Antti, and Huhe Harnud. 2005. Acoustical comparison of the monophthong systems in Finnish, Mongolian and Udmurt. *Journal of the International Phonetic Association* 35: 59–71. [CrossRef]

Jilka, Matthias, Volha Anufryk, Henrike Baumotte, Natalie Lewandowska, Giuseppina Rota, and Susanne Reiterer. 2008. Assessing individual talent in second language production and perception. In *New Sounds 2007: Proceedings of the 5th International Symposium on the Acquisition of Second Language Speech*. Edited by Andréia Schurt Rauber, Michael Alan Watkins and Barbara O. Baptista. Florianópolis: Federal University of Santa Catarina, pp. 243–58.

Johnson, Keith, and Molly Babel. 2010. On the perceptual basis of distinctive features: Evidence from the perception of fricatives by Dutch and English speakers. *Journal of Phonetics* 38: 127–36. [CrossRef]

Jung, Miji, and Sungmi Kwon. 2010. A study on the phonetic discrimination and acquisition ability of Korean language learners. *Malsoriwa Eumseong Gwahak [Phonetics and Speech Sciences]* 2: 23–32.

Jurgec, Peter. 2005. Formant frequencies of standard Slovene vowels. *Govor* 22: 127–44.

Kagaya, Ryohei. 1974. A fiberscopic and acoustic study of the Korean stops, affricates and fricatives. *Journal of Phonetics* 2: 161–80. [CrossRef]

Kang, Yoonjung. 2014. Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics* 45: 76–90. [CrossRef]

Kartushina, Natalia, and Ulrich Hans Frauenfelder. 2013. On the role of L1 speech production in L2 perception: Evidence from Spanish learners of French. In *Proceedings of the 14th Annual Conference of the International Speech Communication Association (INTERSPEECH-2013), Lyon, France, 25–29 August 2013*. Edited by Frédéric Bimbot, Christophe Cerisara, Cécile Fougeron, Guillaume Gravier, Lori Lamel, François Pellegrino and Pascal Perrier. BAIXAS: ISCA Archive, pp. 2118–22.

Kartushina, Natalia, and Ulrich Hans Frauenfelder. 2014. On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in Psychology* 5: 1246. [CrossRef] [PubMed]

Keating, Patricia A. 1984. Phonetic and phonological representation of stop consonant voicing. *Language* 60: 286–319. [CrossRef]

Kiliç, Mehmet Akif, Fatih Öğüt, Gürsel Dursun, Erdoğan Okur, Ilhami Yildirim, and Raşit Midilli. 2004. The effect of vowels on voice perturbation measures. *Journal of Voice* 18: 318–24. [CrossRef] [PubMed]

Kilman, Lisa, Adriana Zekveld, Mathias Hällgren, and Jerker Rönnberg. 2014. The influence of non-native language proficiency on speech perception performance. *Frontiers in Psychology* 5: 651. [CrossRef]

Kim-Renaud, Young-Key. 1974. Korean Consonantal Phonology. Ph.D. dissertation, University of Hawai'i, Honolulu, HI, USA.

Kim-Renaud, Young-Key. 2012. Modern Korean. In *The Languages of Japan and Korea*. Edited by Nicolas Tranter. Abingdon: Routledge, pp. 123–67. [CrossRef]

Ko, Insung. 2009. The merger of *ey* /e/ and *ay* /ɛ / of Seoul Korean. *Eoneo Jeongbo [Language Information]* 10: 73–89.

Kuhl, Patricia K., and Paul Iverson. 1995. Linguistic experience and the "perceptual magnet effect". In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Edited by Winifred Strange. Baltimore: York Press, pp. 121–54.

Kuhl, Patricia K., Barbara T. Conboy, Sharon Coffey-Corina, Denise Padden, Maritza Rivera-Gaxiola, and Tobey Nelson. 2008. Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B* 363: 979–1000. [CrossRef]

Kuhl, Patricia K., Karen A. Williams, Francisco Lacerda, Kenneth N. Stevens, and Björn Lindblom. 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255: 606–8. [CrossRef]

Kwon, Sungmi. 2013. A study on the phonetic sensitivity in the acquisition of non-native language speech sound. *Saegugeo Gyoyuk [Korean Education]* 97: 209–32. [CrossRef]

Labov, William, Sharon Ash, and Charles Boberg. 2006. *The Atlas of North American English: Phonetics, Phonology and Sound Change*. Berlin: Mouton de Gruyter. [CrossRef]

Ladefoged, Peter. 1999. American English. In *Handbook of the International Phonetic Association*. Edited by the International Phonetic Association. Cambridge: Cambridge University Press, pp. 41–44.

Lee, Hyun Bok. 1993. Korean. *Journal of the International Phonetic Association* 23: 28–31. [CrossRef]

Lee, Wai-Sum, and Eric Zee. 2003. Standard Chinese (Beijing). *Journal of the International Phonetic Association* 33: 109–12. [CrossRef]

Levy, Erika S., and Winifred Strange. 2008. Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics* 36: 141–57. [CrossRef]

Maddieson, Ian. 1984. *Patterns of Sounds*. Cambridge: Cambridge University Press.

Major, Roy C. 2001. *Foreign Accent: The Ontogeny and Phylogeny of Second Language Phonology*. Mahwah: Lawrence Erlbaum Associates.

Major, Roy C. 2008. Transfer in second language phonology: A review. In *Phonology and Second Language Acquisition*. Edited by Jette G. Hansen Edwards and Mary L. Zampini. Amsterdam: John Benjamins Publishing, pp. 63–94. [CrossRef]

Martínez-Celdrán, Eugenio, Ana Ma Fernández-Planas, and Josefina Carrera-Sabaté. 2003. Castilian Spanish. *Journal of the International Phonetic Association* 33: 255–59. [CrossRef]

Milovanov, Riia, Minna Huotilainen, Vesa Välimäki, Paulo A. A. Esquef, and Mari Tervaniemi. 2008. Musical aptitude and second language pronunciation skills in school-aged children: Neural and behavioral evidence. *Brain Research* 1194: 81–89. [CrossRef] [PubMed]

Milovanov, Riia, Päivi Pietilä, Mari Tervaniemi, and Paulo A. A. Esquef. 2010. Foreign language pronunciation skills and musical aptitude: A study of Finnish adults with higher education. *Learning and Individual Differences* 20: 56–60. [CrossRef]

Munro, Miles. 2008. CVC syllables for investigating the phonetic sensitivity of Mandarin and English speakers. *Behavior Research Methods* 40: 147–53. [CrossRef]

Myers, Emily B. 2014. Emergence of category-level sensitivities in non-native speech sound learning. *Frontiers in Neuroscience* 8: 238. [CrossRef]

Odlin, Terence. 1989. *Language Transfer: Cross-Linguistic Influence in Language Learning*. Cambridge: Cambridge University Press.

Öğüt, Fatih, Mehmet Akif Kiliç, Erkan Zeki Engin, and Raşit Midilli. 2006. Voice onset times for Turkish stop consonants. *Speech Communication* 48: 1094–99. [CrossRef]

Oh, Eunjin. 2011. Effects of speaker gender on voice onset time in Korean stops. *Journal of Phonetics* 39: 59–67. [CrossRef]

Park, Hansang. 2002a. The time courses of F1 and F2 as a descriptor of phonation types. *Eoneohak [Linguistics]* 33: 87–108.

Park, Hansang. 2002b. Temporal and Spectral Characteristics of Korean Phonation Types. Ph.D. dissertation, University of Texas at Austin, Austin, TX, USA.

Perrachione, Tyler K., Jiyeon Lee, Louisa Y. Y. Ha, and Patrick C. M. Wong. 2011. Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *Journal of the Acoustical Society of America* 130: 461–72. [CrossRef] [PubMed]

Piske, Thorsten. 2008. Phonetic awareness, phonetic sensitivity and the second language learner. In *Encyclopedia of Language and Education*, 2nd ed. Edited by Jasone Cenoz and Nancy H. Hornberger. Boston: Springer, vol. 6 (Knowledge about Language), pp. 155–66. [CrossRef]

Polka, Linda, and Janet F. Werker. 1994. Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance* 20: 421–35. [CrossRef] [PubMed]

R Development Core Team. 2020. R: A Language and Environment for Statistical Computing. Version 4.0.2. Available online: http://www.r-project.org (accessed on 22 June 2020).

Raizada, Rajeev D. S., Feng-Ming Tsao, Huei-Mei Liu, and Patricia K. Kuhl. 2010. Quantifying the adequacy of neural representations for a cross-language phonetic discrimination task: Prediction of individual differences. *Cerebral Cortex* 20: 1–12. [CrossRef]

Roach, Peter. 2004. British English: Received Pronunciation. *Journal of the International Phonetic Association* 34: 239–45. [CrossRef]

Robinson, Peter. 2001. Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition. *Second Language Research* 17: 368–92. [CrossRef]

Schertz, Jessamyn, Taehong Cho, Andrew Lotto, and Natasha Warner. 2015. Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics* 52: 183–204. [CrossRef]

Schertz, Jessamyn, Taehong Cho, Andrew J. Lotto, and Natasha Warner. 2016. Individual differences in perceptual adaptability of foreign sound categories. *Attention, Perception, and Psychophysics* 78: 355–67. [CrossRef]

Schmidt, Anna Marie. 2007. Cross-language consonant identification: English and Korean. In *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*. Edited by Ocke-Schwen Bohn and Murray J. Munro. Amsterdam: John Benjamins Publishing, pp. 185–200. [CrossRef]

Sebastián-Gallés, Núria, Carles Soriano-Mas, Cristina Baus, Begoña Díaz, Volker Ressel, Christophe Pallier, Albert Costa, and Jesús Pujol. 2012. Neuroanatomical markers of individual differences in native and non-native vowel perception. *Journal of Neurolinguistics* 25: 150–62. [CrossRef]

Seoul National University Language Education Institute. 2010. *Active Korean 1*. Seoul: Munjin Media.

Shahidi, A. H., and Rahim Aman. 2011. An acoustical study of English plosives in word initial position produced by Malays. *3L: The Southeast Asian Journal of English Language Studies* 17: 23–33.

Silva, David J. 2006a. Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology* 23: 287–308. [CrossRef]

Silva, David J. 2006b. Variation in voice onset time for Korean stops: A case for recent sound change. *Korean Linguistics* 13: 1–16. [CrossRef]

Skehan, Peter. 1989. *Individual Differences in Second-Language Learning*. London: Edward Arnold.

Skehan, Peter. 1991. Individual differences in second language learning. *Studies in Second Language Acquisition* 13: 275–98. [CrossRef]

Skehan, Peter. 1998. *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.

Skehan, Peter. 2002. Theorizing and updating aptitude. In *Individual Differences and Instructed Language Learning*. Edited by Peter Robinson. Amsterdam: John Benjamins Publishing, pp. 69–93. [CrossRef]

Sohn, Ho-Min. 1999. *The Korean Language*. Cambridge: Cambridge University Press.

Strange, Winifred, Andrea Weber, Erika S. Levy, Valeriy Shafiro, Miwako Hisagi, and Kanae Nishi. 2007. Acoustic variability within and across German, French and American English vowels: Phonetic context effects. *Journal of the Acoustical Society of America* 122: 1111–29. [CrossRef] [PubMed]

Strange, Winifred. 2011. Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics* 39: 456–66. [CrossRef]

Suomi, Kari, Juhani Toivanen, and Riikka Ylitalo. 2008. *Finnish Sound Structure: Phonetics, Phonology, Phonotactics and Prosody*. Oulu: Oulu University Press.

Tyler, Michael D., Catherine T. Best, Alice Faber, and Andrea G. Levitt. 2014. Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica* 71: 4–21. [CrossRef]

Verhoeven, Ludo, and Anne Vermeer. 2002. Communicative competence and personality dimensions in first and second language learners. *Applied Psycholinguistics* 23: 361–474. [CrossRef]

Vokic, Gabriela. 2010. L1 allophones in L2 speech production: The case of English learners of Spanish. *Hispania* 93: 430–52.

Wanrooij, Karin, Paola Escudero, and Maartje E. J. Raijmakers. 2013. What do listeners learn from exposure to a vowel distribution? An analysis of listening strategies in distributional learning. *Journal of Phonetics* 41: 307–19. [CrossRef]

Werker, Janet F. 1994. Cross-language speech perception: Development change does not involve loss. In *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*. Edited by Judith C. Goodman and Howard C. Nusbaum. Cambridge: MIT Press, pp. 93–120.

Werker, Janet F. 1995. Age-related changes in cross-language speech perception: Standing at the crossroads. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Edited by Winifred Strange. Baltimore: York Press, pp. 155–70.

Werker, Janet F., and Judith E. Pegg. 1992. Infant speech perception and phonological acquisition. In *Phonological Development: Models, Research, Implications*. Edited by Charles A. Ferguson, Lise Menn and Carol Stoel-Gammon. Timonium: York Press, pp. 285–311.

Werker, Janet F., and Richard C. Tees. 1984. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7: 49–63. [CrossRef]

Werker, Janet F., and Suzanne Curtin. 2005. PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development* 1: 197–234. [CrossRef]

White, Lydia. 2003. *Second Language Acquisition and Universal Grammar*. Cambridge: Cambridge University Press.

Yang, Byunggon. 1992. An acoustical study of Korean monophthongs produced by male and female speakers. *Journal of the Acoustical Society of America* 91: 2280–83. [CrossRef] [PubMed]

Yoon, Kyuchul, and Soonok Kim. 2015. A comparative study on the male and female vowel formants of the Korean Corpus of Spontaneous Speech. *Malsoriwa Eumseong Gwahak [Phonetics and Speech Sciences]* 7: 131–38. [CrossRef]

Yoon, Tae-Jin, and Yoonjung Kang. 2014. Monophthong analysis on a large-scale speech corpus of read-style Korean. *Malsoriwa Eumseong Gwahak [Phonetics and Speech Sciences]* 6: 139–45. [CrossRef]

languages

Article

Perceived Phonological Overlap in Second-Language Categories: The Acquisition of English /r/ and /l/ by Japanese Native Listeners

Michael D. Tyler

School of Psychology and the MARCS Institute for Brain, Behavior and Development, Western Sydney University, Penrith, NSW 2751, Australia; m.tyler@westernsydney.edu.au

Abstract: Japanese learners of English can acquire /r/ and /l/, but discrimination accuracy rarely reaches native speaker levels. How do L2 learners develop phonological categories to acquire a vocabulary when they cannot reliably tell them apart? This study aimed to test the possibility that learners establish new L2 categories but perceive phonological overlap between them when they perceive an L2 phone. That is, they perceive it to be an instance of more than one of their L2 phonological categories. If so, improvements in discrimination accuracy with L2 experience should correspond to a reduction in overlap. Japanese native speakers differing in English L2 immersion, and native English speakers, completed a *forced category goodness rating task*, where they rated the goodness of fit of an auditory stimulus to an English phonological category label. The auditory stimuli were 10 steps of a synthetic /r/–/l/ continuum, plus /w/ and /j/, and the category labels were L, R, W, and Y. Less experienced Japanese participants rated steps at the /l/-end of the continuum as equally good versions of /l/ and /r/, but steps at the /r/-end were rated as better versions of /r/ than /l/. For those with more than 2 years of immersion, there was a separation of goodness ratings at both ends of the continuum, but the separation was smaller than it was for the native English speakers. Thus, L2 listeners appear to perceive a phonological overlap between /r/ and /l/. Their performance on the task also accounted for their responses on /r/–/l/ identification and AXB discrimination tasks. As perceived phonological overlap appears to improve with immersion experience, assessing category overlap may be useful for tracking L2 phonological development.

Keywords: Perceptual Assimilation Model; second language speech learning; English /r/ and /l/; Japanese; English as a second language; categorical perception; speech perception

Citation: Tyler, Michael D. 2021. Perceived Phonological Overlap in Second-Language Categories: The Acquisition of English /r/ and /l/ by Japanese Native Listeners. *Languages* 6: 4. https://doi.org/10.3390/languages6010004

Received: 1 September 2020
Accepted: 22 December 2020
Published: 28 December 2020

1. Introduction

Adult second language (L2) learners almost invariably speak with a recognizable foreign accent (Flege et al. 1995; Flege et al. 1999). Less obvious for the casual observer is that they are also likely to have difficulty discriminating certain pairs of phonologically contrasting phones in the target language—that is, they also *hear* with an accent (Jenkins et al. 1995). Research into cross-language speech perception by naïve listeners has shown that attunement to the native language affects the discrimination of pairs of phonologically contrasting non-native phones (e.g., Best et al. 2001; Polka 1995; Tyler et al. 2014b; Werker and Logan 1985), often resulting in poor discrimination when both non-native phones are perceived as the same native phonological category.

For learners of an L2, the question is whether and to what extent they are able to overcome their perceptual accent and acquire new phonological categories. Discrimination of initially difficult contrasts, such as English /r/–/l/ for Japanese native listeners, can improve with naturalistic exposure (MacKain et al. 1981) and laboratory training (Bradlow et al. 1999; Bradlow et al. 1997; Lively et al. 1993; Lively et al. 1992; Lively et al. 1994; Logan et al. 1991; Shinohara and Iverson 2018). For example, in Logan et al. (1991), learners identified minimal-pair words containing /r/ or /l/ (e.g., *rake* or *lake*) and were given corrective feedback for incorrect responses. After 15 training sessions of 40 min each, learners' identification showed a small but significant improvement. The results of the series of experiments showed that: (a) the learning generalized to novel talkers and tokens; (b) the perceptual training resulted in improved production, and; (c) improvements in both perception and production were still evident when learners were tested months later. However, in spite of the improvements from high variability training, the learners' performance did not reach the same level of accuracy as that of native speakers of English.

The fact that learners were able to identify words according to whether they contained /r/ or /l/ suggests that they had some sensitivity to the phonetic characteristics that define English /r/ versus /l/. To establish separate lexical entries for /r/–/l/ minimal pairs (e.g., *rock* and *lock*), it could be argued that they must have developed separate phonological categories for /r/ and /l/. If this is the case, it remains to be explained how they might have learned a phonological distinction without being able to discriminate it to the same level as a native speaker. The purpose of this paper is to propose a learning scenario that may account for these observations. L2 learners may, under certain learning circumstances, develop L2 phonological categories that correspond to similar sets of phonetic properties. That is, when they encounter an L2 phone, they may perceive a phonological overlap, such that the L2 phone is consistent with more than one L2 phonological category. This proposal will be tested using perception of English /r/ and /l/ by native speakers of Japanese, which has a long history of investigation. As this idea emerges from recent developments in the Perceptual Assimilation Model (PAM, Best 1994a, 1995), it complements the theoretical framework of its extension to L2 learning, the Perceptual Assimilation Model of Second Language Speech Learning (PAM-L2, Best and Tyler 2007). Thus, this chapter will present up-to-date reviews of both PAM and PAM-L2, and these will be followed by a review of Japanese listeners' perception of /r/ and /l/, and an outline of the present study.

1.1. The Perceptual Assimilation Model

The Perceptual Assimilation Model (PAM) was devised to account for the influence of first language (L1) attunement, both on infants' discrimination of native and non-native contrasts (Best 1994a; Best and McRoberts 2003; Tyler et al. 2014b) and on adults' discrimination of previously unheard non-native contrasts (PAM, Best 1994b; Best 1995; Best et al. 2001; Best et al. 1988; Faris et al. 2018; Tyler et al. 2014a). Following the ecological theory of perception (e.g., Gibson 1966, 1979), PAM proposes that articulatory gestures are perceived directly. Phonological categories are the result of perceptual learning, where the perceiver tunes into the higher-order invariant properties that define the category across a range phonetic variability. While they are abstract, in the sense of being coarser grained than specific articulatory movements, or the acoustic structure that corresponds to those movements, phonological categories are perceptual units rather than mental constructs (Goldstein and Fowler 2003). According to PAM, discrimination accuracy for non-native phones depends on whether and how they are *assimilated* to the native phonological system. When adults encounter non-native contrasts, their natively tuned perception may sometimes help perception (e.g., when the non-native phones are assimilated to two different native categories) and it may sometimes hinder perception (e.g., when both non-native phones are assimilated to the same native category).

An individual non-native phone can be perceptually assimilated as either *categorized* or *uncategorized*, or not assimilated to speech (Best 1995). Categorized phones are perceived as good, acceptable, or deviant versions of a native phonological category. Uncategorized phones are those that are perceived as speech, but not as any one particular native category, and non-assimilable phones are perceived as non-speech. For example, Zulu clicks were often perceived by native English speakers as finger snaps or claps (Best et al. 1988).

The perceptual learning that shapes phonological categories is driven not only by detecting the higher-order phonetic properties that define category membership, but also those set a category apart from other phonological categories in the system (see Best 1994b, pp. 261–62). PAM predicts relative discrimination accuracy for pairs of phonologically contrasting non-native phones according to how each one is assimilated to the native phonological space. That is, discrimination accuracy for never-before-heard non-native phones depends, to a large extent, on whether the listener detects a *native* phonological distinction between the non-native phones. Consider, first, those contrasts where both non-native phones are assimilated to a native category. If each phone is assimilated to a different category (a *two-category* assimilation), then a native phonological distinction will be detected, and discrimination will be excellent. When both contrasting non-native phones are assimilated to the same native category, there is no native phonological distinction to detect and discrimination depends on whether the listener detects a difference in phonetic goodness of fit to the same native category.[1] If there is a goodness difference (a *category-goodness* assimilation), then discrimination will be moderate to very good, depending on the magnitude of the perceived difference, and if there is no goodness difference (a *single-category* assimilation), then discrimination will be poor. Thus, for discrimination accuracy, PAM predicts: two category > category goodness > single category. This main PAM prediction has been confirmed in studies on non-native consonant (Antoniou et al. 2012; Best et al. 2001) and vowel perception (Tyler et al. 2014a).

Turning to contrasts involving uncategorized phones, Best (1995) suggested that when one phone is uncategorized and the other is categorized (an *uncategorized-categorized* assimilation), discrimination should be very good. When both are uncategorized (an *uncategorized-uncategorized* assimilation), discrimination should vary depending on their phonetic proximity to each other. Faris et al. (2016) expanded on this description by describing the different ways that uncategorized non-native phones might be assimilated to the native phonological space. On each trial of a categorization task, Egyptian-Arabic listeners assigned a native orthographic label to an Australian-English vowel (in a /hVbə/ context, where V denotes the target vowel) and then rated its goodness-of-fit. To rule out random responding, Faris et al. tested whether each label was selected above chance for each vowel. An uncategorized native phone was deemed to be *focalized* if only one label was selected above chance (but below the categorization threshold, usually 50% or 70%, Antoniou et al. 2013; Bundgaard-Nielsen et al. 2011b), *clustered* if more than one label was selected above chance, or *dispersed* if no label was selected above chance. It is important to note that, by definition, listeners recognize weak similarity to native phonological categories when non-native phones are perceived as focalized or clustered, but no native phonological similarity when they are perceived as dispersed.

These expanded assimilation types lead to new predictions for uncategorized-categorized and uncategorized-uncategorized assimilations. Faris et al. (2018) suggested that the discrimination of contrasts involving focalized and clustered assimilations should vary according to the overlap in the set of categories that are consistent with one non-native phone versus the set that is consistent with the other phone. This is known as *perceived phonological overlap* (see also, Bohn et al. 2011). For example, if both non-native phones were clustered, and they were weakly consistent with the same set of native phonological categories, then they would be completely overlapping. Discrimination for completely overlapping contrasts should be less accurate than for contrasts that are only partially overlapping, and most accurate for non-overlapping contrasts (i.e., those where unique sets of labels are chosen for each non-native phone). To test this prediction, Australian-English listeners completed categorization with goodness rating and AXB discrimination tasks with Danish vowels. In AXB discrimination, participants were presented with three different vowel (V) tokens (in a /hVbə/ context) and asked to indicate whether the middle element (X) was the same as the first (A) or third (B) element. While Faris et al. did not observe any completely overlapping contrasts, they demonstrated that non-overlapping contrasts were discriminated more accurately than partially overlapping contrasts. This shows that naïve listeners are influenced by perceptual assimilation to the native phonological space even when a non-native phone is perceived as weakly consistent with one or more native

[1] See Tyler (2021) for a discussion of the different sources of information that might be available for the discrimination of non-native contrasts.

phonological categories. The purpose of this study is to demonstrate how perceived phonological overlap might account for speech perception in learners who have acquired new L2 phonological categories, which is the focus of PAM's extension to L2 speech learning, PAM-L2.

1.2. PAM-L2

The Perceptual Assimilation Model of Second Language Speech Learning (PAM-L2, Best and Tyler 2007) was devised for predicting the likelihood of new category formation when a learner acquires an L2. Taking the naïve perceiver described by PAM as its starting point, learners are assumed to accommodate phonological categories from all of their languages in a common interlanguage phonological system. Phonological categories may be shared between the L1 and L2 (and subsequently learned languages), or the learner may establish new L2 categories. Importantly, and in contrast to the Speech Learning Model (Flege 1995, 2003), PAM-L2 proposed that learners could maintain common L1-L2 phonological categories with language-specific L1 and L2 phonetic categories, in a similar way that allophones of a phoneme might be thought to correspond to a single phonological category. For example, an early-sequential Greek-English bilingual's common L1-L2 phonological /p/ category could incorporate language-specific phonetic variants for long-lag aspirated English [$\mathrm{p^h}$] and short-lag unaspirated Greek [p] (Antoniou et al. 2012).

PAM-L2's predictions for new L2 phonological category formation are made on the basis of PAM contrast assimilation types. To illustrate the PAM-L2 principles, Best and Tyler (2007) outlined a number of hypothetical scenarios involving a previously naïve listener acquiring an L2 in an immersion context. In the case of a two-category assimilation, the learner comes already equipped with the ability to detect the phonological difference between the non-native phonemes, through attunement to the L1. Discrimination of two-category contrasts is predicted to be excellent at the beginning of L2 acquisition and the learner would develop a common L1/L2 phonological category for each of the non-native phonemes. No further learning would be required for that contrast, but L2 perception would be more efficient if they developed new phonetic categories for the L2 pronunciations of their common L1/L2 phonological categories. For single-category assimilations, the learner is unlikely to establish a new phonological category for either phone and discrimination will remain poor. In fact, both L2 phonemes are likely to be incorporated into the same L1-L2 phonological category and contrasting words in the L2 that employ those phonemes should remain homophonous for the L2 learner. For uncategorized-categorized assimilations, new L2 phonetic and phonological categories are likely to be established for the uncategorized phone, with the likelihood being higher for non-overlapping and partially overlapping assimilations than for completely overlapping assimilations (Tyler 2019).

Perhaps the most interesting case is category-goodness assimilation. According to PAM-L2, the learner is likely to develop L2 phonetic and phonological categories for the more deviant phone of the contrast. Best and Tyler (2007) speculated that the learner would first establish a new phonetic category for the more deviant phone. Initially, the deviant phone would simply be a phonetic variant of the common L1-L2 phonological category, but as the learner came to recognize that the phonetic difference between the phones signaled an L2 phonological contrast, a new L2 phonological category would emerge for the newly developed phonetic category. This new L2 phonological category would support the development of an L2 vocabulary that maintains a phonological distinction between minimally contrasting words.

Best and Tyler (2007) suggested that this perceptual learning might occur fairly early in the learning process for adult L2 acquisition. Learners with 6–12 months of immersion experience were considered to be "experienced". An increasing vocabulary was seen as a limiting influence on perceptual learning, but more recent work has shown that vocabulary size may assist learners in establishing L2 categories (Bundgaard-Nielsen et al. 2012; Bundgaard-Nielsen et al. 2011a, 2011b). It is possible that vocabulary expansion constrains perceptual learning in the case of a single category assimilation, but facilitates perceptual learning when there is a newly established L2 phonological category (i.e., for category goodness assimilations and those involving an uncategorized phone). For example, once new L2 phonological categories are established, learners could use lexically guided perceptual retuning (McQueen et al. 2012; Norris et al. 2003; Reinisch et al. 2013) to accommodate to the phonetic properties of the newly acquired category.

The scenarios outlined in Best and Tyler (2007) assumed an idealized situation where an adult learner with no previous L2 experience is immersed in the L2 environment and where L2 input is entirely through the spoken medium. However, the majority of learners do not acquire an L2 solely from spoken language—the L2 is often acquired first in a formal learning situation using both written and spoken language, and this may occur in the learner's country of origin from a teacher who may speak the L2 with a foreign accent. While category formation could occur from the bottom up in an immersion scenario, the classroom learning environment may expose learners to information about an L2 phonological contrast before they have had the opportunity to attune to the phonetic properties necessary to discriminate it. When the L2 has a phonographic writing system, the most likely source of information would be from orthographic representations of words (see Tyler 2019, for a discussion of how orthography might influence category acquisition in the classroom).

The classroom learning environment would not change the PAM-L2 predictions for contrasts that were initially two-category assimilations, and the uncategorized phone of an uncategorized-categorized assimilation may even benefit from more rapid acquisition under those circumstances (Tyler 2019). However, for those contrasts where both L2 phones are assimilated to the same L1 category, single-category and category-goodness assimilations, the language learning environment could have profound effects on attunement. In fact, this situation may result in a new type of scenario that was not considered in Best and Tyler (2007). Let us reconsider the case of a category-goodness assimilation under those circumstances. The L1 category to which both L2 phonemes are assimilated would form a common L1-L2 phonological category with the more acceptable L2 phoneme, as in the immersion case. For the more deviant phone, rather than discovering a new L2 phonological category on the basis of attunement to articulatory-phonetic information, as was proposed for the immersion context, the learners could possibly discover the L2 phonological contrast via other sources (e.g., via orthography when it has unambiguous grapheme-phoneme correspondences). Since phonological categories are perceptual units, according to PAM-L2, the learners would need to establish a new L2 phonological category for the more deviant phone to acquire L2 words that preserved the phonological contrast. If they had not yet tuned into the phonetic differences that signal the phonological contrast in the L2, then the new L2 phonological category would correspond to a similar set of phonetic properties as the common L1-L2 category. In that situation learners may continue to have difficulty discriminating the L2 contrast, not because the two phones are assimilated to the same phonological category, but because both phones are perceived as being instances of the same two phonological categories. Just as Faris et al. (2016, 2018) have shown that a pair of non-native phones may fall in a region of phonetic space that corresponds to the same set of L1 phonological categories (i.e., a completely overlapping uncategorized-uncategorized assimilation), it is proposed here that a pair of L2 phones might be perceived as consistent with the same two L2 phonological categories.

The aim of this study is to test whether L2 learners who have acquired English in the classroom prior to immersion perceive phonological overlap between L2 phonological categories for contrasting L2 phones that were likely to have been initially perceived as category-goodness assimilations. Furthermore, with immersion experience, the categories should start to separate, such that learners who have recently arrived in an English-speaking country should exhibit greater category overlap than L2 users who have been living in the L2 environment for a long period (MacKain et al. 1981).

1.3. Perception of English /r/ and /l/ by Japanese Native Listeners

The learner group for the present study will be Japanese native speakers and the contrast to be tested will be the English /r/–/l/ contrast. Although the /r/–/l/ contrast was originally thought to be a single-category assimilation (Best and Strange 1992), there is now widespread agreement that it is a category-goodness assimilation (Aoyama et al. 2004; Guion et al. 2000; Hattori and Iverson 2009), with /l/ being rated as a more acceptable version of the Japanese /ɾ/ category than is English /r/. Using this contrast allows the results to be interpreted in light of the rich history of investigations into that listener group/contrast combination.

One of the earliest investigations into Japanese identification and discrimination of /r/ and /l/ was conducted by Goto (1971). He established that Japanese native speakers learning English had difficulty identifying and discriminating minimal-pair words, such as *play* and *pray* and concluded that /r/–/l/ pronunciation difficulties were likely to be perceptual in origin. Miyawaki et al. (1975) had participants discriminate steps on a synthetic /r/–/l/ continuum, in which F1 and F2 were held constant and F3 varied. Unlike native English speakers, Japanese native speakers living in Japan with at least 10 years of formal English language training did not show a categorical peak in discrimination, suggesting that they did not perceive a categorical boundary between /r/ and /l/. Interestingly, the Japanese listeners performed similarly to native English listeners when they were presented with a non-speech continuum that contained only the F3 transition (F1 and F2 values were set to zero). Thus, the Japanese listeners were able to detect the differences in frequencies along the F3 continuum, but when the same acoustic patterns were presented in a speech context, they failed to discriminate them.

In contrast to the findings of Miyawaki et al. (1975), which tested discrimination only, Mochizuki (1981) found that a group of Japanese listeners residing in the USA split the continuum into two separate categories in an /r/–/l/ identification task. This naturally led to the hypothesis that Japanese native speakers can acquire the /r/–/l/ distinction given sufficient naturalistic exposure. To test this, MacKain et al. (1981) directly compared the identification and discrimination of an /r/–/l/ continuum by Japanese native speakers who differed in their exposure to English conversation. Both groups were living in the USA at the time of testing; one group had received extensive English conversation training from a native speaker of US English whereas the other group had received little or no native English conversation training. To optimize the possibility that the less experienced group might discriminate the stimuli, MacKain et al. enriched the /r/–/l/ continuum by providing multiple redundant cues. Whereas Miyawaki et al. held F1 and F2 constant while varying F3, the stimuli of MacKain et al. varied along all three dimensions. In spite of the redundant cues, the less experienced group showed no evidence of categorical perception and their discrimination was close to chance. The more experienced group, on the other hand, perceived the stimuli categorically, and in a similar way to native US-English speakers, and although their discrimination was less accurate than the native speakers, the shape of the response function was similar. They concluded that it was indeed possible for Japanese native speakers to acquire categorical perception of English /r/ and /l/ that approximates that of native speakers.

An individual-differences approach to identification of /r/ and /l/ by Japanese native speakers was undertaken by Hattori and Iverson (2009). Their 36 participants varied in age (19 to 48 years), amount of formal English instruction (7 to 25 years), and the amount of time spent living in an English-speaking country (1 month to 13 years, with a median of 3 months). Participants completed two types of identification task. One was an /r/–/l/ identification task, in which participants heard minimal-pair English words, such as *rock* and *lock*, and indicated whether they began with /r/ or /l/. There was a wide range of accuracy, from close to chance to 100% correct, and a mean of 67% correct. The other identification task was a "bilingual" task, with consonant-vowel syllables formed from the combination of five vowels and three consonants (English /r/, English /l/, and Japanese /ɾ/). Participants were asked to indicate whether the first sound was R, L, or Japanese

R. The identification of /r/ was quite accurate, at 82% correct, and it was almost never confused with /ɾ/ (2%). Identification was less accurate for /l/ (58% correct), with errors split evenly between /r/ and /ɾ/. Japanese /ɾ/ was identified reasonably accurately (77% correct), with most confusions occurring with /l/ (16%). Clearly, the fewest confusions occurred between /r/ and /ɾ/, and the most confusion occurred for /l/ with both /r/ and /ɾ/. While the authors suggested that /l/ appears to be assimilated to Japanese /ɾ/, they did not observe a correlation between /r/–/l/ identification accuracy and the degree of confusion between /ɾ/ and /r/. That is, those who performed poorly on /r/–/l/ identification appear not to have done so because they assimilated both /r/ and /l/ to Japanese /ɾ/. Instead, they suggested that that the learners had an /r/category that was not optimally tuned to English. It is possible that those results could be explained by perceived phonological overlap. The results as a whole are consistent with the idea being proposed here that the participants had developed a new L2 phonological category for /r/, with a corresponding English [ɹ] phonetic category, and that /l/ had assimilated to /ɾ/ to form a common L1-L2 phonological category, with language-specific [l] and [ɾ] phonetic variants.[2] Variability in identification for /r/, /l/, and /ɾ/ could be explained by the different patterns of phonetic and phonological overlap between the categories.

1.4. The Present Study

The aim of this study is to test the idea that learners who had acquired English in a formal learning setting may perceive phonological overlap when they encounter L2 phones, and that the overlap decreases with immersion experience in an L2 environment. Faris et al. (2018) determined overlap in L1 cross-language speech perception using a categorization task with goodness rating. Across the sample, if the same set of categories was selected above chance for two non-native vowels then the contrast was deemed to be overlapping. While this approach has been shown to provide a reliable indication of overlap in cross-language speech perception, particularly for vowels where categorization is more variable than it is for consonants, there is an assumption built into the categorization task that an individual only perceives one phonological category in the stimulus. If the participant perceives the stimulus as consistent with more than one category, then the categorization task may only provide an imperfect approximation of the amount of overlap. For example, if the stimulus is clearly more acceptable as one category than another, then the participant may only ever select the best-fitting category, and the task would fail to reveal perceived phonological overlap.

For the purposes of this study, a task is required that can identify perceived phonological overlap, for a given L2 phone, and that is capable of detecting differences in the amount of overlap between groups. This can be achieved by eliminating the categorization stage and simply rating the goodness of fit of each stimulus to a phonological category that is provided on each trial—a *forced category goodness rating task*. For example, participants would be presented with an English category label, e.g., “R as in ROCK”, and an auditory stimulus, and they would be asked to rate the goodness of fit of the auditory stimulus to the given category. That task will be used in the present study to assess category overlap in Japanese learners of English with more versus less experience in an immersion environment, and in a native English control group.

The stimuli will be the 10-step rock-lock continuum, developed at Haskins Laboratories (Best and Strange 1992; Hallé et al. 1999; MacKain et al. 1981), that uses multiple redundant cues for F1, F2, and F3. The /w/ and /j/ end points of the wock-yock continuum will also be included as control stimuli, as participants are expected to indicate that those stimuli are not similar to either /r/ or /l/. Participants will rate the goodness of fit of each stimulus to four English phonological categories: /l/, /r/, /w/, and /j/. Results from across the continuum will give insight into the internal structure of the phonological category, at least along one axis of acoustic-phonetic variability, and they will also pro-

[2] Note that square brackets are used here to denote a phonetic category rather than a specific phone.

vide a link to previous studies that have used those stimuli. To further assist with such comparisons, participants will complete discrimination and /r/–/l/ identification tasks in addition to the forced category goodness rating task. In line with those previous studies, native English listeners should show categorical perception of stimuli from the rock-lock continuum in both identification and discrimination, with progressively less categorical perception observed for more experienced, then less experienced Japanese listeners.

In the forced category goodness rating task, native English listeners should rate [w], [j], and the [ɹ] and [l] ends of the continuum as good examples only for the corresponding phonological category (e.g., /w/ for [w]), and as low on each of the other three categories. Ratings for the continuum steps should vary in a similar way to the identification task. The more and less experienced Japanese participants should both rate [w] and [j] as good only for the corresponding phonological category, and low on each of the other three categories. If participants perceive phonological overlap for stimuli along the rock-lock continuum, then forced goodness ratings will be above the lowest rating for more than one phonological category. It is anticipated that the ratings as /r/ and /l/ will be high for both continuum end points, but that the difference between /r/ and /l/ goodness ratings for the same stimulus (e.g., [ɹ]) will be greater for the more than less experienced groups. That is, the overlap should be smaller for more experienced than less experienced Japanese learners of English.

2. Materials and Methods

2.1. Participants

Japanese native speakers were recruited from Sydney, Australia, via word-of-mouth, noticeboard advertisements at local universities, and a Japanese-language electronic bulletin board service targeted at expatriate Japanese people living in Sydney. The aim was to recruit two samples of Japanese native speakers: (1) migrants who had been living in Australia for a long period, and; (2) recent arrivals. Fifty-five Japanese native speakers were tested. Data were discarded for four participants who were immersed in an English-speaking country at the age 16 or younger, and for one participant who was raised in Hong Kong. To establish a clear difference in length of residence (LOR) between the more experienced and less experienced groups, data were retained for participants who had been living in Australia for a minimum of 2 years, or for 3 months or less. Data for 13 participants with LORs ranging from 5 to 19 months were therefore discarded (M = 0.89 years).

The final sample of Japanese native speakers consisted of 19 more experienced English users (16 females, M_{age} = 38 years, Age Range: 21 to 59 years, M_{LOR} = 8 years, LOR Range: 2 to 27 years) and 18 less experienced English users (13 females, M_{age} = 25 years, Age Range: 20 to 35 years, M_{LOR} = 8 weeks, LOR Range: 1 to 13 weeks). The participants were given a small payment for their participation in the study.

The participants were asked to report any languages that they learned outside the home (i.e., in a formal education context) and the age at which they began to learn them. For the more experienced group, all participants began to learn English in Japan between 9 and 13 years of age (M = 12 years, SD = 0.91 years). One participant did not report the number of years of English study, but the remainder reported between 6 and 15 years (M = 9 years, SD = 2.41 years). For the less experienced group, the participants began to learn English in Japan between 5 and 13 years of age (M = 11 years, SD = 1.9 years). Two participants did not report the number of years of English study, with the remainder completing between 6 and 15 years (M = 10 years, SD = 2.42 years).

The Australian-English native speakers were recruited from the graduate and undergraduate student population at Western Sydney University, Australia, who received course credit for participation. There were 16 participants (14 females, M_{age} = 21 years, Age Range: 18 to 35 years). Data for an additional four participants were collected but discarded due to childhood acquisition of a language other than English (n = 1), self-reported history of a language disorder (n = 2), or brain injury (n = 1).

2.2. Stimuli and Apparatus

Participants were presented with the 10-step /rak/–/lak/ (*rock-lock*) continuum that was first used in MacKain et al. (1981), and the endpoints of the /wak/–/jak/ (*wock-yock*) continuum from Best and Strange (1992); see also (Hallé et al. 1999). The first consonant and vowel portions of the stimuli were generated with the OVE-IIIc cascade formant synthesizer at Haskins Laboratories, and the /k/ was appended to the synthesized syllables. See the original articles for additional stimulus details, including F1, F2, and F3 parameters. A questionnaire was used to collect basic demographic information, and information about the participants' language learning history. The experiment was run on a MacBook laptop running Psyscope X B50 (http://psy.ck.sissa.it/). Participants listened to stimuli through Koss UR-20 headphones set at a comfortable listening level.

2.3. Procedure

Participants completed the forced category goodness rating task, followed by /r/–/l/ identification, and AXB discrimination. The language background questionnaire was completed at the end of the session.

Forced category goodness rating. Participants were instructed that on each trial they would hear a syllable in their headphones and that their task would be to decide how similar the first sound of the syllable was to one of four English consonant categories, presented on screen (R as in ROCK, L as in LOCK, W as in WOCK, or Y as in YOCK). They were asked to rate the similarity on a 7-point scale, using the numbers on the computer keyboard, where 1 indicated that it was highly similar to the given category, 4 was somewhat similar, and 7 was highly dissimilar.[3] Participants were encouraged to try to use the entire scale from 1–7 across the experiment, and they were instructed not to reflect too long on their response. If they did not respond within 4 s, the trial was aborted, and a message instructed them to respond more quickly. Missed trials were repeated later in the list to ensure that each participant provided a full set of rating data. Each of the 12 stimulus tokens (10 /r/–/l/ continuum steps plus the /w/–/j/ endpoints) were presented five times, each in the context of the four rating categories, resulting in a total of 240 trials. Stimuli were randomized within each of the five blocks. To maintain participant vigilance and to give an opportunity to take a short break, the participant was asked to press the space bar to continue every 20 trials. The task took approximately 20 min to complete.

Identification. The 10 steps of the /rak/–/lak/ continuum were used in the identification task. Participants were instructed to listen to syllables through headphones and indicate whether the first sound was more like "r" as in "rock" or "l" as in "lock". They responded using the D and L keys on the keyboard, which were labeled with "R" and "L", respectively. The letters R and L were also displayed on the left- and right-hand side of the screen. Participants were instructed not to reflect for too long on their response. The trial timed out after 2 s, which was followed by a message on screen to respond more quickly. Missed trials were reinserted at a random point in the remaining trial sequence. Participants were presented with 20 randomly ordered blocks of the 10 steps (200 trials in total) and they pressed the space bar at the end of each block to continue. The test took approximately 10 min to complete.

AXB Discrimination. Following Best and Strange (1992), participants were tested using AXB discrimination. Three tokens were presented sequentially. The first and third were different steps on the continuum and the middle item was identical to the first or the third token. Participants were tested on steps that differed by three along the continuum ([ɹ]-4, 2-5, 3-6, 4-7, 5-8, 6-9, 7-[l]). Stimuli were presented in all four possible AXB trial combinations (AAB, ABB, BAA, BBA). These 28 trials (7 step pairs $\times$ 4 AXB trial combinations) were randomized within blocks, which were presented 5 times. Each step pair was therefore presented 20 times in total across 140 trials. Participants were told that on each trial they

3 Although participants had no difficulty using this scale, the scale is reversed in the results section to assist the reader in interpreting the data patterns. That is, 7 is reported as *highly similar* and 1 is reported as *highly dissimilar*.

would hear three syllables in their headphones one after the other, that the first and third syllables were different, and the second syllable was either the same as the first or the third syllable. They were instructed to indicate whether the second syllable was the same as the first or the third syllable, using the keys 1 and 3 on the computer keyboard, basing their decision only on the first sound in each syllable (i.e., the consonant). The software would not allow the participants to respond until the third syllable had finished playing. Participants were instructed not to reflect for too long on their response. A trial timed out after 2 s and was repeated later in the experiment. It took around 5 min to complete.

3. Results

3.1. Forced Category Goodness Rating

Participants rated the 10 steps of the /r/–/l/ continuum and the endpoints of the /w/–/j/ continuum against the categories /l/, /r/, /w/, and /j/. The results for the three groups are presented in Figure 1. The top-left panel (a) shows the data for less experienced Japanese listeners, the top-right panel (b) for more experienced Japanese listeners, and the bottom panel (c) for Australian-English listeners. The auditory stimulus is plotted on the *x*-axis. Step 1 of the /r/–/l/ continuum is denoted as [ɹ], step 10 is [l], and the other steps are denoted by their step number. The endpoints of the /w/–/j/ continuum are denoted as [w] and [j]. The mean ratings are plotted on the *y*-axis. Thus, the top left point on the English native listener plot represents participants' mean ratings for how well auditory [ɹ] fit the /r/ category, "R"; the bottom left point represents their ratings for auditory [ɹ] to the /l/ category, "L".

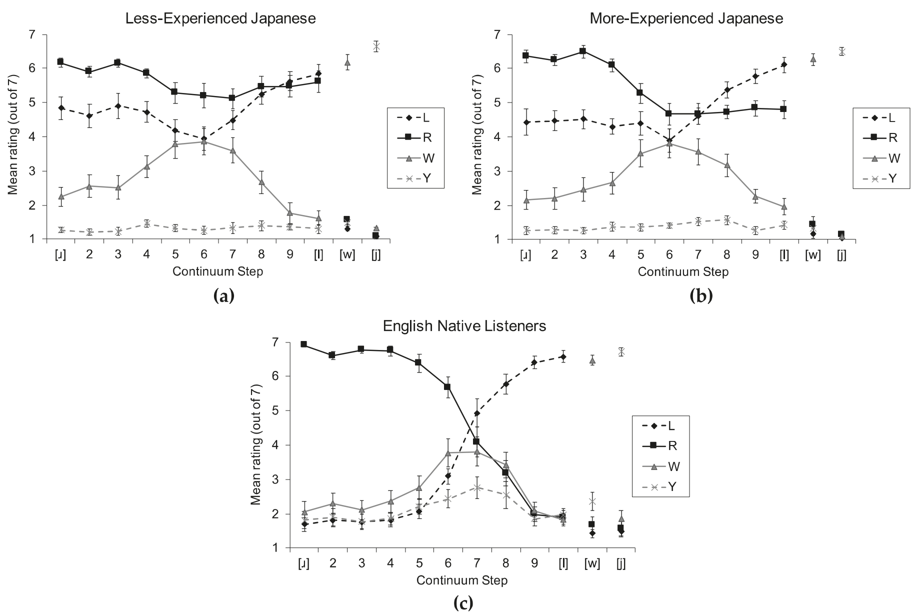


Figure 1. Mean forced goodness ratings for each auditory stimulus, to /l/ (L), /r/ (R), /w/ (W), and /j/ (Y) categories, where 1 is "highly dissimilar" and 7 "highly similar". (**a**). Less experienced Japanese (<3 months immersion); (**b**) More experienced Japanese (>2 years immersion); (**c**) Native English listeners. Error bars represent standard error of the mean.

The Australian-English group results, shown in the bottom panel of Figure 1, show a classic categorical perception pattern for the /r/–/l/ continuum for ratings as /l/ and /r/, with the cross-over point between steps 6 and 7. The phones [w] and [j] were each rated as highly similar to /w/ and /j/, respectively, and highly dissimilar to any other category. It is interesting to note that the ambiguous regions of the /r/–/l/ continuum also appear to have attracted higher acceptability ratings for /w/ and /j/ than the /r/–/l/ endpoints did. To test this, the ratings to /w/ and /j/ for step 7 were compared to those for [l] and [ɹ], respectively, in separate 2 × 2 repeated measures analyses of variance (ANOVAs). For step 7 versus [l], there was a main effect of step, $F(1, 15) = 20.79$, $p < 0.001$, $\eta^2_G = 0.27$,[4] and a significant two-way interaction between step and category, $F(1, 15) = 9.09$, $p < 0.001$, $\eta^2_G = 0.06$, indicating that the rating difference between step 7 and [l] was greater for /w/ than for /j/. Simple effects tests showed that the rating difference between step 7 and [l] was significant for both /w/, $F(1, 15) = 20.92$, $p < 0.001$, $\eta^2_G = 0.38$, and /j/, $F(1, 15) = 8.46$, $p = 0.01$, $\eta^2_G = 0.13$. For step 7 versus [ɹ], there were main effects of step, $F(1, 15) = 26.92$, $p < 0.001$, $\eta^2_G = 0.22$, and category, $F(1, 15) = 10.36$, $p = 0.006$, $\eta^2_G = 0.06$, but no interaction, $F(1, 15) = 3.86$, $p = 0.07$. This suggests that both step 7 and [ɹ] are more /w/-like than /j/-like, and that, overall, the ratings for step 7 are higher than those for [ɹ]. Together, these results indicate that the most ambiguous step was perceived as somewhat /w/-like and weakly /j/-like, in addition to being perceived as somewhat /r/- and /l/-like. The [l] endpoint was not perceived as similar to /w/ or /j/, and it appears that the listeners may have perceived the [ɹ] endpoint to have some similarity to /w/ but not /j/.

An initial comparison of the pattern of results for the two Japanese groups, in the top-left and top-right panels of Figure 1, suggests that there may be a difference in their ratings of the /r/-/l/ continuum. Considering first the endpoints, the less experienced group appear to rate [l] as both an acceptable /l/ and an acceptable /r/, whereas [ɹ] appears to be rated as a more acceptable /r/ than /l/. This supports the idea that they perceive a phonological overlap between /r/ and /l/ for both [ɹ] and [l], but at first glance it suggests that they already have a reasonable sensitivity to the difference between /r/ and /l/ at the [ɹ] end of the continuum. It is important to note, however, that the ratings as /r/ across the entire continuum are uniformly high. The separation in ratings for [ɹ] as /r/ and as /l/ may be due to that phone being perceived as a poorer /l/ rather than a more acceptable /r/. The more experienced Japanese group, on the other hand, appear to have rated [ɹ] as a more acceptable /r/ than /l/, and [l] as a more acceptable /l/ than /r/. The lower of the two goodness ratings for both [ɹ] and [l] are around 4—"Somewhat Acceptable". There is also a remarkable similarity between the shape of the response curve for ratings as /l/ for both groups at the /l/-end of the continuum, which is consistent with the idea that they have established a common L1-L2 category for English /l/ and Japanese /ɾ/.

To test these observations, the ratings as /r/ and /l/ for the 10 steps of the /r/–/l/ continuum were subjected to a 2 × (2) × (10) mixed ANOVA. The between-subjects variable was *group* (less experienced vs. more experienced Japanese listeners) and the two within-subjects variables were *category* (/l/ vs. /r/) and *step*. There were main effects of category, $F(1, 35) = 40.47$, $p < 0.001$, $\eta^2_G = 0.09$, and step, $F(9, 315) = 12.17$, $p < 0.001$, $\eta^2_G = 0.09$, and a significant two-way interaction between them, $F(9, 315) = 18.96$, $p < 0.001$, $\eta^2_G = 0.13$. The differential responding by the more experienced and less experienced groups, that can be seen in Figure 1, was confirmed by a significant three-way interaction between category, step, and group, $F(9, 315) = 2.67$, $p = 0.005$, $\eta^2_G = 0.02$. To explore the three-way interaction further, separate two-way ANOVAs were run for each group. There were two-way interactions between category and step for both the more experienced, $F(9, 162) = 14.93$, $p < 0.001$, $\eta^2_G = 0.19$, and the less experienced groups, $F(9, 153) = 5.20$, $p < 0.001$, $\eta^2_G = 0.07$. Another set of two-way ANOVAs was conducted to test whether the groups differed for each category. There was a two-way interaction between step and group for /r/, $F(9, 315) = 2.59$, $p = 0.007$, $\eta^2_G = 0.04$, but not for /l/, $F(9, 315) = 0.68$, $p = 0.73$, suggesting that the differences between the groups can be accounted for by improvements in perception of their /r/ category only. To test for differences in goodness ratings as /r/

[4] Generalized eta squared (η^2_G) is a measure of effect size that is appropriate for mixed designs (Olejnik and Algina 2003). It is compatible with Cohen's (1988) benchmarks for interpreting eta squared (small = 0.01, medium = 0.06, and large = 0.14).

versus /l/ at each continuum step, post-hoc paired *t*-tests were run separately for each group, with a Bonferroni-adjusted alpha level of 0.005. The results are presented in Table 1. For the less experienced group, there are significant differences between ratings as R versus L at steps [ɹ] through 4, and for step 6. For the more experienced group, the ratings are also different for steps [ɹ] through 4. Importantly, and unlike the less experienced group, they are also different for step [l].

Table 1. Post-hoc paired *t*-tests for ratings as /r/ versus /l/ at each continuum step, for less experienced and more experienced Japanese listeners.

Group		**Step**									
		[ɹ]	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**[l]**
Less experienced	$t(17)$	**−3.43**	**−3.84**	**−3.65**	**−4.30**	−2.66	**−4.81**	−2.84	−1.37	0.75	0.71
	p	**0.003**	**0.001**	**0.002**	**<0.001**	0.016	**<0.001**	0.011	0.190	0.462	0.488
More experienced	$t(18)$	**−6.22**	**−4.42**	**−5.58**	**−5.02**	−2.53	−2.40	−0.19	2.07	2.34	**4.11**
	p	**<0.001**	**<0.001**	**<0.001**	**<0.001**	0.021	0.028	0.850	0.054	0.031	**0.001**

Values in boldface are significant at a Bonferroni-adjusted alpha rate of 0.005.

Like the Australian-English listeners, both Japanese groups rated [w] and [j] as highly similar to /w/ and /j/, respectively, and not to any other category. The ratings as /w/ across the /r/–/l/ continuum appear to be similar in the groups. They also appear to have similar ratings as /j/, but with a flatter response than the English native listeners. The ratings as /w/ and /j/ for step 7 and [l], and for step 7 and [ɹ], were compared for the two Japanese groups using separate 2 × (2) × (2) ANOVAs. The between-subjects variable was *group* (less experienced vs. more experienced) and the within-subjects variables were *category* (/w/ vs. /j/) and *step* (step 7 vs. [l] or step 7 vs. [ɹ]). For step 7 versus [l], there were main effects of category, $F(1, 35) = 35.77$, $p < 0.001$, $\eta^2_G = 0.20$, and step, $F(1, 35) = 21.22$, $p < 0.001$, $\eta^2_G = 0.10$, and a significant two-way interaction between them, $F(1, 35) = 15.90$, $p < 0.001$, $\eta^2_G = 0.07$. Crucially, there was no three-way interaction between category, step, and group, $F(1,35) = 0.05$, $p = 0.83$, suggesting that the more and less experienced Japanese listeners responded similarly to each other. The significant two-way interaction was further probed with tests of simple effects, which showed that the differences in ratings for step 7 and [l] were significant for ratings as /w/, $F(1, 35) = 22.08$, $p < 0.001$, $\eta^2_G = 0.03$, but not for ratings as /j/, $F(1, 35) = 1.37$, $p = 0.25$. The results were similar for step 7 versus [ɹ], with the main effects of category, $F(1, 35) = 48.30$, $p < 0.001$, $\eta^2_G = 0.27$, and step, $F(1, 35) = 13.73$, $p = 0.001$, $\eta^2_G = 0.06$, and a significant two-way interaction between them, $F(1, 35) = 4.18$, $p < 0.001$, $\eta^2_G = 0.02$.

3.2. Identification

The mean percent L responses for identification of steps from the /r/–/l/ continuum are presented in Figure 2. The native English listeners show the classic ogive-shaped categorical perception function that was observed in previous studies using the same stimuli (Best and Strange 1992; Hallé et al. 1999; MacKain et al. 1981). The boundary location appears to be between steps 6 and 7, mirroring the results from the forced category goodness rating task. This seems to be closer to the /l/ end of the continuum than the American-English listeners in the previous studies, whose boundary location was between steps 5 and 6. The more experienced Japanese participants appear to have a shallower function than the native English participants and the endpoints of the continuum do not reach 0% or 100% (at 19% and 77%, respectively). The less experienced Japanese participants' function appears to be even shallower than the more experienced participants' and their endpoints are closer to the chance level of 50% ([ɹ] = 32%, [l] = 58%).

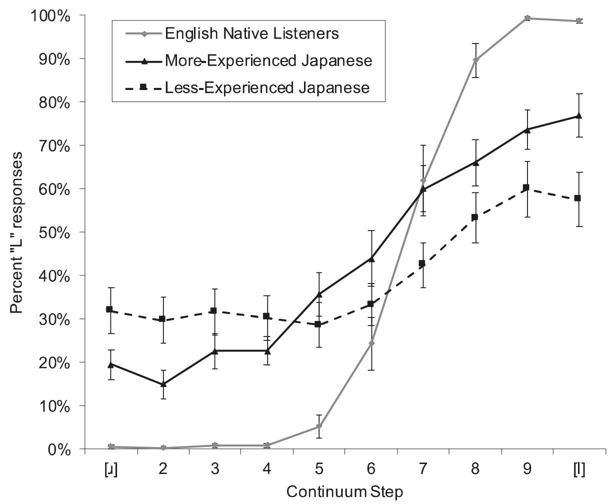


Figure 2. Mean percent "L" responses from the R-L identification task for the three participant groups. Error bars represent standard errors of the mean.

Given the English listeners' ceiling performance for steps [ɹ]-4, 9, and [l], and the corresponding lack of variability, it is not appropriate to use standard parametric tests to compare their performance to the Japanese groups'. It is clear that they perform differently from the more experienced Japanese listeners. Furthermore, given that the shapes of the response curves for most of the Japanese participants in this data set do not appear to follow an ogive-shaped function, the data have not been fit to a cumulative distribution function, as they were in previous studies (Best and Strange 1992; Hallé et al. 1999). A 2 × (10) ANOVA was conducted to test whether the response curves differed for more experienced versus less experienced Japanese listeners. The between-subjects factor was *experience* and the within-subjects factor was *step*. The shape of the response curve was tested using orthogonal polynomial trend contrasts on the step factor—linear, quadratic (one turning point), cubic (two turning points), and quartic (three turning points) (Winer et al. 1991, for contrast coefficients). There was no significant main effect of experience, $F(1, 35) = 0.75$, but there were significant overall linear, $F(1, 35) = 60.39$, $p < 0.001$, quadratic, $F(1, 35) = 7.89$, $p = 0.008$, cubic, $F(1, 35) = 17.76$, $p < 0.001$, and quartic trends, $F(1, 35) = 5.34$, $p = 0.02$. Those significant trend contrasts simply indicate that the curve has a complex shape. The important question is whether there are any significant interactions between the trend contrasts and experience. The only significant interaction was with the linear trend contrast, $F(1, 35) = 7.91$, $p = 0.008$. Therefore, while there is no evidence for a difference in the shape of the Japanese participants' response curve, the significant interaction shows that the more experienced group have a generally steeper function than the less experienced group, as can be seen in Figure 2.

3.3. Discrimination

Mean percent correct responses for discrimination of the seven pairs of steps from the continuum are presented in Figure 3. Again, there are clear differences between the English and Japanese listeners' performance. The English listeners show the classic categorical perception discrimination response, with poorer discrimination for steps that are on the same side of the categorical boundary and more accurate discrimination for steps that cross the category boundary. The Japanese participants, on the other hand, show a clear double peak.

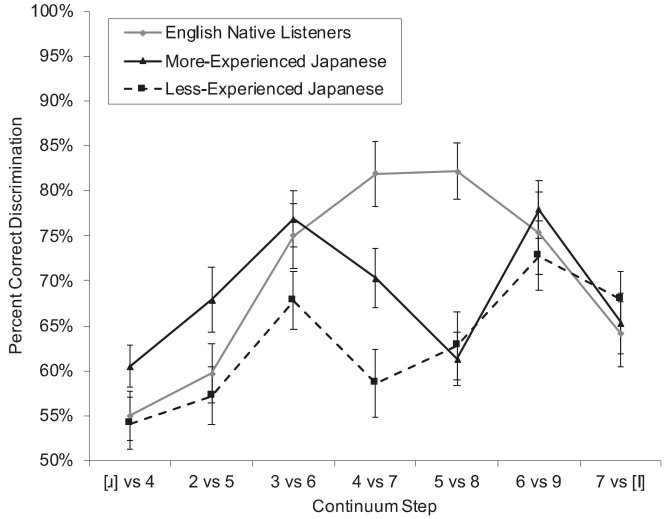


Figure 3. Mean percent correct discrimination of pairs of steps from the /r/–/l/ continuum for the three participant groups. Error bars represent standard errors of the mean.

As the performance of the native speakers is not at ceiling on this task, it is possible to conduct a 3 × (7) ANOVA comparing all three groups. The between-subjects factor was *group*, with two planned contrasts. The *language background* contrast compared the English listeners with the combined results of the two Japanese groups and the *experience* contrast compared the two Japanese groups only. The within-subjects factor was *step pair*, and the shape of the curve was tested again using orthogonal polynomial contrasts. The language background contrast was not significant, $F(1, 50) = 3.47$, but a significant experience contrast showed that the more experienced group performed more accurately, overall, than the less experienced group, $M_{\text{difference}} = 5.53\%$, $SE = 2.74$, $F(1, 50) = 4.08$, $p = 0.049$. There was a significant overall linear trend, $F(1, 50) = 27.48$, $p < 0.001$, reflecting the gradual increase in accuracy, collapsed across all three groups, from the [ɹ]-end to the [l]-end of the continuum. There was no interaction between the linear trend contrast and language background, but it interacted significantly with experience, $F(1, 50) = 4.49$, $p = 0.04$. This can be seen in Figure 3, where the less experienced group appear to be relatively less accurate at the [ɹ]-end than the [l]-end of the continuum, whereas the more experienced group show more similar levels of accuracy at both ends. There was also a significant overall quadratic trend, $F(1, 50) = 33.81$, $p < 0.001$, which interacted with the language background contrast only, $F(1, 50) = 20.90$, $p < 0.001$. This reflects the fact that the English participants' responses show a quadratic shape, whereas the Japanese participants' responses appear to follow a quartic shape. Indeed, there was no significant cubic trend, and no interactions, and no significant overall quartic trend, but there was a significant interaction between the quartic trend and language background, $F(1, 50) = 9.67$, $p = 0.003$.

4. Discussion

The aim of this study was to test whether certain L2 learners, who initially acquired their L2 in a formal learning context, might perceive an L2 phone as an instance of more than one L2 phonological category, and whether that perceived phonological overlap is smaller with longer versus shorter periods of L2 immersion. To test for perceived phonological overlap, participants completed a forced category goodness rating task, where they were presented with an auditory token and rated its goodness of fit to a given English phonological category label, L, R, W, or Y. It was hypothesized that Japanese native speakers who had first been exposed to English in Japan would perceive a phonological overlap between /r/ and /l/, but that the overlap would be smaller for those with a long period of immersion in an English-speaking environment (>2 years) than those with a short period of immersion (<3 months). Native English speakers rated [ɹ] as a highly similar to /r/, but not /l/, /w/, or /j/, and [l] as highly similar to /l/, but not to the other three categories. In contrast, the Japanese native speakers rated [ɹ] as highly similar to /r/, moderately similar to /l/, and dissimilar to /w/ and /j/. The two groups of Japanese speakers differed from each other in their ratings of [l]. The less experienced Japanese group rated [l] as highly similar to *both* /r/ and /l/, whereas the more experienced Japanese group rated [l] as highly similar to /l/ and only moderately similar to /r/. In short, the hypotheses were supported. When the Japanese listeners perceived stimuli along the continuum between /r/ and /l/, they perceived varying degrees of phonological overlap with /r/ and /l/, and the overlap was smaller for the more experienced than the less experienced group. This study is cross-sectional, so it is not possible to conclude definitively that there is a reduction in overlap due to immersion experience, but these results are certainly consistent with that idea.

The general pattern of results observed for the less experienced group was predicted by a learning scenario, proposed here, that had not previously been considered by Best and Tyler (2007) when they outlined PAM-L2. The scenarios presented in Best and Tyler were based on functional monolinguals who were acquiring an L2 in an immersion setting. In contrast, the participants in both Japanese groups had learned English in Japan prior to immersion. As the /r/–/l/ contrast is a category-goodness assimilation for Japanese native speakers (e.g., Guion et al. 2000), the PAM-L2 learning scenario suggests that [l] would initially be perceived as a common L1-L2 phonetic category with Japanese [ɾ], and that Japanese /ɾ/ and English /l/ would form a common L1-L2 phonological category. English [ɹ] would first be perceived as an allophonic variant of the common L1-L2 /ɾ/–/l/ category, and then a new L2-only /r/ phonological category would be established when the learner recognized that the phonetic contrast between /r/–/l/ signaled a phonological distinction. Thus, the non-native category-goodness contrast would become an L2 two-category contrast. Such a situation should have resulted in a data pattern that resembled the native English speakers' results. Since the Japanese native speakers in this study acquired English in a formal learning situation in Japan, it was argued here that they may have needed to establish a new phonological category for /r/ before they had managed to tune in to its phonetic properties. That learning scenario is consistent with the data pattern observed in this study. The participants were able to indicate that steps along the continuum were perceived as similar to /r/, but they rated the same stimuli as also having various degrees of similarity to /l/. The fact that they rated the steps as being dissimilar to /j/ shows that they were not simply indicating that any L2 consonant was similar to /r/. Rather, it is consistent with the idea that they had developed a new phonological category for /r/ that was poorly tuned to the phonetic properties that distinguish it from /l/. In PAM terms, both /r/ and /l/ would be uncategorized clustered assimilations. Rather than becoming a two-category assimilation, /r/–/l/ would be a completely overlapping uncategorized-uncategorized contrast (Faris et al. 2016, 2018).

A close inspection of Figure 1 shows that the response pattern for ratings as /l/ appears to be similar for the less and more experienced groups, with stimuli closer to [ɹ] rated as poorer /l/s than those closer to the [l] end. Indeed, follow-up analyses of the three-way interaction between group, category, and step showed that the difference between the two groups was entirely due to differences in ratings as /r/. The less experienced group gave high ratings as /r/ across the continuum, whereas the more experienced group gave lower ratings as /r/ towards the [l]-end of the continuum. The finding that the ratings as /l/ were unaffected by immersion experience fits with the idea that English /l/ is a common L1-L2 category with Japanese /ɾ/, as an L1 category may be more resistant to change with L2 experience than a new L2-only phonological category. Future research could investigate this further by including Japanese /ɾ/ as a rating category and auditory stimulus, in addition to English /r/ and /l/.

Presenting the /r/–/l/ continuum, rather than isolated tokens of /r/ and /l/, allows us to compare the internal structure of the phonological categories. It is clear from comparing the three groups in Figure 1 that the English native listeners have a much clearer separation of their /r/ and /l/ categories than the Japanese listeners. Steps [ɹ]-4 have phonetic properties that appear to be prototypically /r/ for the English listeners, and those of steps 9-[l] appear to be prototypically /l/. The uniformly high ratings as /r/ across the continuum for the less experienced Japanese group suggests that their L2 English /r/ category may initially cover a broad region of phonetic space, with minimal differences in phonetic goodness of fit. Improvements that have been observed previously after periods of immersion may, therefore, be due more to improvements in category tuning to /r/ than /l/ (e.g., MacKain et al. 1981). Training regimes for improving /r/–/l/ perception in Japanese native listeners have traditionally focused on identifying minimal-pair words (e.g., Bradlow et al. 1997; Hattori and Iverson 2009; Lively et al. 1993; Lively et al. 1992; Lively et al. 1994), that is, on recognizing those phonetic characteristics that define category membership of /r/ versus those that define /l/. If the learner perceives a phonological overlap, and the token is perceived as an equally good example of both /r/ and /l/, then the utility of that training may be limited. The findings of the present study suggest that there may be some benefit in training learners to recognize which tokens should be a good versus poor fit to a category rather than identifying which category they belong to, before transitioning to training regimes that focus on detecting the relational phonetic differences that characterize phonological distinctions (e.g., using a discrimination task).

4.1. Identification and Discrimination

In contrast to the forced category goodness rating task, where participants rated the phonetic goodness of fit to a phonological category that was provided on each trial, the identification task required them to make a forced choice between two phonological categories. In an identification ask, a participant could decide that a given stimulus was /r/, for example, because either it clearly sounded like /r/ or it clearly did *not* sound like /l/. As the forced category goodness rating task provides some insight into the listeners' judgement of the extent to which each continuum step resembled or did not resemble /l/ or /r/, their ratings should be related to their identification response function. As the forced category goodness ratings as /l/ and /r/ are basically mirror images of each other for the English native listeners, they could have made their decision using either criterion, and their identification response curve would have been similar to the one presented in Figure 2. For both Japanese groups, the separation in forced category goodness ratings as /l/ versus /r/ is wider at the [ɹ]-end than the [l]-end of the continuum. This is reflected in the identification accuracy, given that both groups were more accurate at identifying [ɹ] than [l] (see Figure 2). That result is consistent with Hattori and Iverson (2009), and with a recent study where Japanese listeners were more accurate for /r/ than /l/ in a *ranby* versus *lanby* identification task (Kato and Baese-Berk 2020). The authors of that study suggested that the asymmetry was due to a bias towards the category that was more dissimilar to the closest native category. The results of this study complement that conclusion by providing a novel theoretical explanation for the source of the bias. Japanese listeners identify /r/ more accurately than /l/ because the perceived phonological overlap between /r/ and /l/ is less pronounced for /r/ than it is for /l/. It would be interesting to see whether the same pattern of overlap is observed for the more dissimilar phone in category-goodness contrasts from other languages and language groups.

Given that the two groups did not differ on their ratings as /l/ in the forced category goodness rating task, they should have had similar judgements in the identification task about what did and did not sound like /l/. Relative differences between the groups' identification function should be attributable to their divergent ratings of each step as /r/. Indeed, all of the steps were acceptable-sounding versions of /r/ for the less experienced group (Figure 1a) and the shape of their identification function (Figure 2) is strikingly similar to that for their ratings as /l/. The more experienced group appears to have a slightly wider separation than the less experienced group between ratings as /r/ and /l/ for steps [ɹ]-4 (Figure 1b), which is reflected in identification by a lower percentage of "L" responses on steps [ɹ]-4. Ratings as /r/ dropped and remained steady from steps 5-[l] (Figure 1b) and this corresponds to the relatively higher accuracy of the more versus less experienced group on identification of those steps (Figure 2). Thus, the forced category goodness rating task has provided insights into the reason why Japanese listeners do not show categorical perception across an /r/-/l/continuum. Differences in goodness of fit to simultaneously perceived phonological categories modulates their judgements of whether the stimulus is or is not a member of each category.

MacKain et al. (1981) compared identification and discrimination of the same continuum by Japanese native speakers living in the US who had undergone intensive conversation training in English, another group with little or no such training, and native speakers of English. The experienced group in that study showed categorical perception along the rock-lock continuum, with an identification response that did not differ from the native speakers', whereas the inexperienced group showed a fairly flat response function. The identification results in the present study were comparable to MacKain et al. for the less experienced group, but the more experienced group was not similar to the native speakers in this study. One key difference between the studies was the criterion used to select the more experienced group. Whereas MacKain et al. selected participants on the basis of conversational training, here they were selected on the basis of length of residence. Focused conversational training in an immersion situation may have resulted in more native-like L2-learning outcomes than simply residing in an L2 environment. A comparison of the present results with those of MacKain et al. would support Flege's (2009, 2019) contention that access to quality native-speaker input is a more direct predictor of L2 perceptual learning than length of residence.

In discrimination, the English speakers showed a clear peak across the categorical boundary, in line with previous research. There is a clear double peak for both Japanese groups, which is in contrast to the relatively flat distribution observed in MacKain et al. (1981). The identification results do not seem to provide any explanation for why a double peak was observed. For example, the less experienced group showed a peak for step 3 versus step 6, but these were identified similarly. A double peak is an indication that the participants may have perceived a third category in the middle of the continuum. The forced category goodness ratings suggest that there was some degree of /w/ perceived in the middle of the continuum. Although it may seem unlikely that participants would have identified /w/ in the middle of the continuum when the goodness ratings for /w/ were no higher than they were for /r/ or /l/, previous findings of /w/ identification in the middle of other /r/–/l/ continua (Iverson et al. 2003; Mochizuki 1981) make this a plausible explanation. Another possibility is that they perceived a different category in the middle of the continuum (e.g., their native Japanese /ɾ/ or possibly the vowel-consonant sequence /ɯɾ/, see Guion et al. 2000), but as they were not asked to rate the stimuli against other categories, that is a question that would need to be addressed in future research. Discrimination accuracy generally increased for the less experienced group along the continuum from [ɹ] to [l], whereas the relatively more accurate discrimination of the more experienced group was fairly level. This may reflect the more experienced group's greater sensitivity to goodness differences in both /r/ and /l/, whereas the less experienced group may have relied primarily on goodness differences relative to /l/ only.

4.2. Conflicting Findings between Pre-Lexical and Lexical Tasks

As phonological categories are pre-lexical perceptual units for PAM-L2, perceived phonological overlap would be a logical consequence of acquiring a phonological category before sufficient perceptual learning had taken place to differentiate it from other categories in the phonological system. The results of this study, and of Kato and Baese-Berk (2020), are consistent with that account. However, in a study examining the time course of L2 spoken word recognition, Cutler et al. (2006) observed an asymmetry that appears to be the reverse of the one observed here in pre-lexical tasks. Japanese and English native speakers heard an instruction to click on one of four objects presented on the screen while an eye tracker monitored their eye movements. On critical trials, one picture depicted an object containing /r/ (e.g., *writer*), another containing /l/ (e.g., *lighthouse*), and there were two non-competitor pictures containing neither /r/ nor /l/. The /r/-/l/ word pairs were chosen so that the onsets overlapped phonologically, such that the participants would need wait for disambiguating information (e.g., the /h/ of *lighthouse*) if they were unable to tell /r/ and /l/ apart. When the word containing /r/ was the target, the Japanese participants took longer to settle their gaze on the correct object than the English native speakers did, suggesting that they were unable to disambiguate the words on the basis of /r/ or /l/. However, when the /l/-word was the target, they settled on the correct picture early, at the same point in time as the English native speakers. Thus, there was an asymmetry in word recognition, such that they were apparently more efficient at recognizing words beginning with /l/ than those beginning with /r/ (see Weber and Cutler 2004, for similar results on Dutch listeners' recognition of English words containing /ɛ/ and /æ/). Given that perceived phonological overlap in this study was smaller for /r/ than /l/, and /r/ is identified more accurately than /l/ (Kato and Baese-Berk 2020), it is surprising that spoken word recognition should show an asymmetry in the opposite direction (see Amengual 2016; Darcy et al. 2013 for other examples of a mismatch between performance pre-lexical and lexical tasks). Cutler et al. explained their results in terms of lexical processing, rather than perception of phonological categories, which may account for the difference. They suggested that the Japanese listeners had established lexical entries that preserved the /r/–/l/ phonological distinction, even though they could not reliably discriminate the contrast, and provided two possible explanations for the asymmetry. One possibility (also suggested by Weber and Cutler 2004) is that when /r/ is included in a lexical entry, it does not receive any bottom-up activation from speech, and nor does it inhibit (or is it inhibited by) the activation of other words as they compete for selection as the most likely word candidate. Activation of the word containing /r/ (and inhibition of other competitors) would only proceed via input that matched its other phonemes. The second possibility is that both /l/ and /r/ words contain the L1 /ɾ/ category in their lexical entries. Words containing /l/ would be activated when a reasonable sounding /ɾ/ is encountered and those containing /r/ would be activated when encountering a poorer match. By that account, the asymmetry arose because /r/ would never be perceived as a reasonable match for /ɾ/, but /l/ could be perceived as a poorer match for /r/. Thus, /l/ would only ever contact words in the lexicon containing /l/, but there is a reasonable probability that /r/ would contact words containing both /l/ and /r/.

Darcy et al. (2013) also concluded that lexical encoding was responsible for the asymmetry. They showed that, in spite of accurate discrimination of L2 Japanese singleton-geminate contrasts or German front-back vowel contrasts, lexical decision performance for L2 learners was poorer than it was for native speakers, particularly for nonword items. They also observed an asymmetry in lexical decision. The stimulus words contained either a more or less native-like L2 phoneme and the nonwords were created by swapping the target phoneme with the other member of the pair (e.g., the German word for 'honey', *Honig* /honɪç/, became the nonword **Hönig* /hønɪç/). Accuracy for words was higher when the category was more versus less nativelike, and accuracy for minimal-pair nonwords was higher when the category was less versus more native-like. Similar to Cutler et al. (2006), Darcy et al. concluded that lexical coding for the less native-like category is fuzzy, and that it encodes the goodness of fit to the dominant L1 category. Interestingly, advanced German learners did not show the asymmetry, which suggests that lexical encoding can improve with L2 experience.

PAM-L2 (Best and Tyler 2007) may provide a slightly different perspective to the conclusions of Cutler et al. (2006) and Darcy et al. (2013). For PAM (Best 1995) and PAM-L2, phonological categories are perceptual, and they are the result of attunement to the higher-order phonetic properties that are relevant both for recognizing words and for telling them apart from other words in the language. As /l/ is initially perceived by Japanese native listeners as a good instance of /ɾ/, their existing L1 /ɾ/ category would be used for acquiring any English words containing /l/ (a common L1-L2 category), and a new L2-only phonological category would be established for English /r/. In spoken word recognition, then, words containing /l/ would benefit from an existing L1 category that is already integrated into processes of lexical competition. In contrast, it may take some time for words containing a new L2 category (i.e., /r/) to establish inhibitory connections that would reduce the activation of competitor words (as may have eventually occurred for the advanced German learners in Darcy et al. 2013). Thus, when the Japanese native listeners in Cutler et al. (2006) perceived /l/, their native lexical competition processes would have inhibited activation of competitor words, including the one containing /r/. They may also have perceived /r/ pre-lexically, but without the benefit of inhibitory connections to other words, the /r/ competitor word would have been inhibited by the word containing /l/. For target words containing /r/, both the /r/- and /l/-words would be activated, but the poorer fit to the /l/ category would limit the activation of the /l/-word competitor. Thus, both candidates remained activated until disambiguating information was encountered. Clearly, more research needs to be done to tease apart the pre-lexical and lexical influences on L2 speech perception.

4.3. Methodological Considerations

The forced category goodness rating task was devised for this study because categorization with goodness rating might underestimate the perceived phonological overlap. To illustrate, Japanese participants rated step 8 as having various degrees of similarity to /l/, /r/, and /j/, but the ratings for /j/ were lower than for the other two categories. Had they completed a categorization test first, they may not have selected "Y" at all because the other three categories are clearly a better fit. A categorization task was not included for comparison here because the session was already quite long. Nevertheless, it is clear that the forced category goodness rating task is capable of detecting category overlap, and that category overlap was observed for both the native and non-native listeners.

The success of the forced category goodness rating task at detecting perceived phonological overlap raises the question of whether it should be adopted in favor of the standard categorization with goodness rating task. Indeed, Faris et al. (2018) suggested that it might be necessary to reconsider the use of arbitrary categorization criteria and the forced category goodness rating task removes the necessity of specifying a threshold for categorization. It may also solve a problem with the categorization of vowels; some participants have difficulty using the keyword labels for identifying vowels, particularly for a language like English, where some of the grapheme–phoneme correspondences are ambiguous. Faris et al. familiarized participants with the 18 English vowel labels, using English vowel stimuli and providing feedback, but they found that up to a quarter of the participants had difficulty selecting the correct label. While this could mean that those participants had difficulty categorizing native vowels, a more likely possibility is that they had poor phonological awareness, and that affected their ability to perform well on that metalinguistic task. Forced category goodness rating might alleviate that problem because the participants are provided with the category against which to judge the auditory stimuli and they do not need to search for the category that corresponds to the vowel that they heard. However, one clear limitation of the forced category goodness rating is that it is much more labor-intensive than categorization. In the case of Faris et al., participants would need to have rated 32 Danish vowels multiple times against the 18 English vowel categories. This would have resulted in thousands of trials. This is not to say that a forced category goodness rating task should be avoided. If it proved to provide a more accurate estimation of perceptual assimilation, then researchers would need to devote the time necessary to collecting those data.

If the forced category goodness rating task was adopted as a test of perceptual assimilation, then there would no longer be arbitrary thresholds for determining whether a non-native phone was assimilated as categorized to the native phonological system. Instead, a non-native phone could be deemed to be categorized as a given native phonological category if the mean rating of the stimulus to that category was significantly above the lowest possible rating (e.g., 1 out of 7, where the 1 is defined as "no similarity"). Expanding on Faris et al. (2016), non-native phones would be categorized as focalized if only one category had a non-negligible rating, categorized as clustered if more than one category had a non-negligible rating, or uncategorized as dispersed if no category had a non-negligible rating.

In spite of the potential for this task to support future theoretical advances in research on cross-language and second-language speech perception, careful comparisons need to be made between the data obtained from current methods and the forced category goodness rating task before suggesting a change to standard research protocols. For example, the English native speakers in this study did not give the lowest possible score for [ɹ] as /w/ (2.05 out of 7), even though they clearly perceived that stimulus as belong to their /r/ category ([ɹ] as /r/ was rated at 6.90 out of 7). This suggests that ratings at the lower end of the scale may reflect phonetic similarity rather than phonological category membership. It may nevertheless be necessary to ask participants to make a decision about category membership rather than relying solely on a goodness-of-fit judgement. In fact, Tyler (2021) has identified four different sources of information that non-native listeners could use to discriminate non-native phones. Any new method for assessing perceptual assimilation would need to be capable of assessing listeners' sensitivity to any information available to them for discriminating given non-native contrast. Until careful methodological studies have been completed comparing different approaches to categorization, studies testing PAM/PAM-L2 predictions should continue to use categorization with goodness rating, giving participants the opportunity to select from among all possible vowel or consonant categories (Bundgaard-Nielsen et al. 2011a, 2011b; Faris et al. 2016).

5. Conclusions

In an ideal learning situation, adults would tune in to the phonetic and phonological properties of an L2 prior to establishing a large L2 vocabulary. However, this is not the way that L2s are generally learned. Classroom-based learning is more common and it provides opportunities to learn about phonological distinctions before attuning to the phonetic properties that define phonological categories and distinguish them from each other. It was argued here that such a situation may give rise to perceived phonological overlap between L2 categories. The results of a forced category goodness rating task showed that Japanese native speakers who first acquired English in the classroom perceived varying degrees of phonological overlap between English /r/ and /l/ when they encountered either category in speech. The overlap was smaller for those with more than two years of immersion experience, as compared to those with less than three months, suggesting that learners continue to attune to the phonological distinction with appropriate input. Assessment of perceived phonological overlap in L2 learners may help with tracking phonological development and with tailoring perceptual training to those contrasts where lexically guided perceptual retuning is most likely to be effective (see Tyler 2019, for a discussion of how PAM-L2 might apply to classroom foreign language acquisition). Future research should investigate category overlap using natural stimuli, and test whether discrimination accuracy for /r/ and /l/ can be predicted by the degree of phonological overlap between the L2 /r/ and /l/ categories.

Funding: This research was funded by the Australian Research Council, grant number DP0880913.

Institutional Review Board Statement: The protocol was approved by the University of Western Sydney Human Research Ethics Committee (Approval 07/040).

Informed Consent Statement: All subjects gave their informed consent for inclusion before they participated in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available as informed consent was not obtained for publishing the data at the time of testing.

Acknowledgments: Thank you to Rikke Bundgaard-Nielsen for assistance with participant recruitment, Louise de Beuzeville for assistance with manuscript preparation, Kikuko Nakamura, Atomi Ohama, Susan Wijngaarden, and Mark Antoniou for research assistance, and two anonymous reviewers for helpful comments that greatly improved the manuscript. The Haskins continua were kindly provided by Pierre Hallé and Catherine Best.

Conflicts of Interest: The author declares no conflict of interest.

References

Amengual, Mark. 2016. The perception of language-specific phonetic categories does not guarantee accurate phonological representations in the lexicon of early bilinguals. *Applied Psycholinguistics* 37: 1221–51. [CrossRef]

Antoniou, Mark, Catherine T. Best, and Michael D. Tyler. 2013. Focusing the lens of language experience: Perception of Ma'di stops by Greek and English bilinguals and monolinguals. *Journal of the Acoustical Society of America* 133: 2397–411. [CrossRef] [PubMed]

Antoniou, Mark, Michael D. Tyler, and Catherine T. Best. 2012. Two ways to listen: Do L2-dominant bilinguals perceive stop voicing according to language mode? *Journal of Phonetics* 40: 582–94. [CrossRef]

Aoyama, Katsura, James E. Flege, Susan G. Guion, Reiko Akahane-Yamada, and Tsuneo Yamada. 2004. Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics* 32: 233–50. [CrossRef]

Best, Catherine T. 1994a. The emergence of native-language phonological influences in infants: A perceptual assimilation model. In *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*. Edited by Judith C. Goodman and Howard C. Nusbaum. Cambridge: MIT Press, pp. 167–244.

Best, Catherine T. 1994b. Learning to perceive the sound pattern of English. In *Advances in Infancy Research*. Edited by C. Rovee-Collier and Lewis P. Lipsitt. Norwood: Ablex, pp. 217–304.

Best, Catherine T. 1995. A direct realist view of cross-language speech perception. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Edited by Winifred Strange. Baltimore: York Press, pp. 171–204.

Best, Catherine T., and Gerald W. McRoberts. 2003. Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech* 46: 183–216. [CrossRef]

Best, Catherine T., Gerald W. McRoberts, and Elizabeth Goodell. 2001. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America* 109: 775–94. [CrossRef]

Best, Catherine T., Gerald W. McRoberts, and Nomathemba M. Sithole. 1988. Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance* 14: 345–60. [CrossRef]

Best, Catherine T., and Winifred Strange. 1992. Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics* 20: 305–30. [CrossRef]

Best, Catherine T., and Michael D. Tyler. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In *Second Language Speech Learning: The Role of Language Experience in Speech Perception and Production*. Edited by Murray J. Munro and Ocke-Schwen Bohn. Amsterdam: John Benjamins, pp. 13–34.

Bohn, Ocke-Schwen, Catherine T. Best, Cinzia Avesani, and Mario Vayra. 2011. Perceiving through the lens of native phonetics: Italian and Danish listenerś perception of English consonant contrasts. In *Proceedings of the 17th International Congress of Phonetic Sciences, Hong Kong, China, August 17–21*. Edited by Wai Sum Lee and Eric Zee. Hong Kong: City University of Hong Kong, pp. 336–39.

Bradlow, Ann R., Reiko Akahane-Yamada, David B. Pisoni, and Yoh'ichi Tohkura. 1999. Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics* 61: 977–85.

Bradlow, Ann R., David B. Pisoni, Reiko Akahane-Yamada, and Yoh'ichi Tohkura. 1997. Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America* 101: 2299–310. [CrossRef]

Bundgaard-Nielsen, Rikke L., Catherine T. Best, Christian Kroos, and Michael D. Tyler. 2012. Second language learners' vocabulary expansion is associated with improved second language vowel intelligibility. *Applied Psycholinguistics* 33: 643–64. [CrossRef]

Bundgaard-Nielsen, Rikke L., Catherine T. Best, and Michael D. Tyler. 2011a. Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition* 33: 433–61. [CrossRef]

Bundgaard-Nielsen, Rikke L., Catherine T. Best, and Michael D. Tyler. 2011b. Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Applied Psycholinguistics* 32: 51–67. [CrossRef]

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale: Lawrence Erlbaum Associates.

Cutler, Anne, Andrea Weber, and Takashi Otake. 2006. Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics* 34: 269–84. [CrossRef]

Darcy, Isabelle, Danielle Daidone, and Chisato Kojima. 2013. Asymmetric lexical access and fuzzy lexical representations in second language learners. *The Mental Lexicon* 8: 372–420. [CrossRef]

Faris, Mona M., Catherine T. Best, and Michael D. Tyler. 2016. An examination of the different ways that non-native phones may be perceptually assimilated as uncategorized. *Journal of the Acoustical Society of America* 139: EL1–EL5. [CrossRef] [PubMed]

Faris, Mona M., Catherine T. Best, and Michael D. Tyler. 2018. Discrimination of uncategorised non-native vowel contrasts is modulated by perceived overlap with native phonological categories. *Journal of Phonetics* 70: 1–19. [CrossRef]

Flege, James E. 1995. Second language speech learning: Theory, findings, and problems. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Edited by Winifred Strange. Baltimore: York Press, pp. 233–76.

Flege, James E. 2003. Assessing constraints on second-language segmental production and perception. In *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*. Edited by Niels O. Schiller and Antje S. Meyer. Berlin: Mouton de Gruyter, pp. 319–55.

Flege, James E. 2009. Give input a chance! In *Input Matters in SLA*. Edited by Thorsten Piske and Martha Young-Scholten. Bristol: Multilingual Matters, pp. 175–90.

Flege, James E. 2019. A non-critical period for second-language learning. In *A Sound Approach to Language Matters—In Honor of Ocke-Schwen Bohn*. Edited by Anne M. Nyvad, Michaela Hejná, Anders Højen, Anna B. Jespersen and Mette H. Sørensen. Aarhus: Department of English, School of Communication & Culture, Aarhus University, pp. 501–41.

Flege, James E., Murray J. Munro, and Ian R. A. MacKay. 1995. Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America* 97: 3125–34. [CrossRef]

Flege, James Emil, Grace H. Yeni-Komshian, and Serena Liu. 1999. Age constraints on second-language acquisition. *Journal of Memory & Language* 41: 78–104.

Gibson, James J. 1966. *The Senses Considered as Perceptual Systems*. Boston: Houghton-Mifflin.

Gibson, James J. 1979. *The Ecological Approach to Visual Perception*. Boston: Houghton-Mifflin.

Goldstein, Louis, and Carol A. Fowler. 2003. Articulatory phonology: A phonology for public language use. In *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*. Edited by Niels O. Schiller and Antje S. Meyer. Berlin: Mouton de Gruyter, pp. 159–207.

Goto, Hiromu. 1971. Auditory perception by normal Japanese adults of the sounds "l" and "r". *Neuropsychologia* 9: 317–23. [CrossRef]

Guion, Susan G., James E. Flege, Reiko Akahane-Yamada, and Jesica C. Pruitt. 2000. An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America* 107: 2711–724. [CrossRef] [PubMed]

Hallé, Pierre A., Catherine T. Best, and Andrea Levitt. 1999. Phonetic vs. phonological influences on French listeners' perception of American English approximants. *Journal of Phonetics* 27: 281–306. [CrossRef]

Hattori, Kota, and Paul Iverson. 2009. English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *Journal of the Acoustical Society of America* 125: 469–79. [CrossRef] [PubMed]

Iverson, Paul, Patricia K. Kuhl, Reiko Akahane-Yamada, Eugen Diesch, Yoh'ichi Tohkura, Andreas Kettermann, and Claudia Siebert. 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87: B47–B57. [CrossRef]

Jenkins, James J., Winifred Strange, and Linda Polka. 1995. Not everyone can tell a "rock" from a "lock": Assessing individual differences in speech perception. In *Assessing Individual Differences in Human Behavior: New Concepts, Methods, and Findings*. Edited by David J. Lubinski and René V. Dawis. Palo Alto: Davies-Black Publishing, pp. 297–325.

Kato, Misaki, and Melissa M. Baese-Berk. 2020. The effect of input prompts on the relationship between perception and production of non-native sounds. *Journal of Phonetics* 79: 100964. [CrossRef]

Lively, Scott E., John S. Logan, and David B. Pisoni. 1993. Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America* 94: 1242–55. [CrossRef]

Lively, Scott E., David B. Pisoni, and John S. Logan. 1992. Some effects of training Japanese listeners to identify English /r/ and /l/. In *Speech Perception, Production and Linguistic Structure*. Edited by Yoh'ichi Tohkura, Eric Vatikiotis-Bateson and Yoshinori Sagisaka. Washington: IOS Press, pp. 175–96.

Lively, Scott E., David B. Pisoni, Reiko A. Yamada, Yoh'ichi Tohkura, and Tsuneo Yamada. 1994. Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America* 96: 2076–87. [CrossRef]

Logan, John S., Scott E. Lively, and David B. Pisoni. 1991. Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America* 89: 874–86. [CrossRef]

MacKain, Kristine S., Catherine T. Best, and Winifred Strange. 1981. Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics* 2: 369–90. [CrossRef]

McQueen, James M., Michael D. Tyler, and Anne Cutler. 2012. Lexical retuning of children's speech perception: Evidence for knowledge about words' component sounds. *Language Learning and Development* 8: 317–39. [CrossRef]

Miyawaki, Kuniko, James J. Jenkins, Winifred Strange, Alvin M. Liberman, Robert Verbrugge, and Osamu Fujimura. 1975. An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics* 18: 331–40.

Mochizuki, Michiko. 1981. The identification of /r/ and /l/ in natural and synthesized speech. *Journal of Phonetics* 9: 283–303. [CrossRef]

Norris, Dennis, James M. McQueen, and Anne Cutler. 2003. Perceptual learning in speech. *Cognitive Psychology* 47: 204–38. [CrossRef]

Olejnik, Stephen, and James Algina. 2003. Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods* 8: 434–47. [CrossRef]

Polka, Linda. 1995. Linguistic influences in adult perception of non-native vowel contrasts. *Journal of the Acoustical Society of America* 97: 1286–96. [CrossRef]

Reinisch, Eva, Andrea Weber, and Holger Mitterer. 2013. Listeners retune phoneme categories across languages. *Journal of Experimental Psychology: Human Perception and Performance* 39: 75–86. [CrossRef]

Shinohara, Yasuaki, and Paul Iverson. 2018. High variability identification and discrimination training for Japanese speakers learning English /r/–/l/. *Journal of Phonetics* 66: 242–51. [CrossRef]

Tyler, Michael D. 2019. PAM-L2 and phonological category acquisition in the foreign language classroom. In *A Sound Approach to Language Matters—In Honor of Ocke-Schwen Bohn*. Edited by Anne M. Nyvad, Michaela Hejná, Anders Højen, Anna B. Jespersen and Mette H. Sørensen. Aarhus: Dept. of English, School of Communication & Culture, Aarhus University, pp. 607–30.

Tyler, Michael D. 2021. Phonetic and phonological influences on the discrimination of non-native phones. In *Second Language Speech Learning: Theoretical and Empirical Progress*. Edited by Ratree Wayland. Cambridge: Cambridge University Press, pp. 157–74.

Tyler, Michael D., Catherine T. Best, Alice Faber, and Andrea G. Levitt. 2014a. Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica* 71: 4–21. [CrossRef]

Tyler, Michael D., Catherine T. Best, Louis M. Goldstein, and Mark Antoniou. 2014b. Investigating the role of articulatory organs and perceptual assimilation in infants' discrimination of native and non-native fricative place contrasts. *Developmental Psychobiology* 56: 210–27. [CrossRef]

Weber, Andrea, and Anne Cutler. 2004. Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language* 50: 1–25. [CrossRef]

Werker, Janet F., and John S. Logan. 1985. Cross-language evidence for three factors in speech perception. *Perception & Psychophysics* 37: 35–44.

Winer, Benjamin J., Donald R. Brown, and Kenneth M. Michels. 1991. *Statistical Principles in Experimental Design*, 3rd ed. New York: McGraw-Hill.

languages

MDPI

Article

The Role of Acoustic Similarity and Non-Native Categorisation in Predicting Non-Native Discrimination: Brazilian Portuguese Vowels by English vs. Spanish Listeners

Jaydene Elvin [1,2,*], Daniel Williams [2,3], Jason A. Shaw [4,5], Catherine T. Best [4,6] and Paola Escudero [2,4]

1 Department of Linguistics, California State University, Fresno, CA 93740, USA
2 ARC Centre of Excellence for the Dynamics of Language, Australian National University, Canberra, ACT 0200, Australia; daniel.williams@uni-potsdam.de (D.W.); paola.escudero@westernsydney.edu.au (P.E.)
3 Linguistics Department, University of Potsdam, 14469 Potsdam, Germany
4 The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Sydney, NSW 2751, Australia; jason.shaw@yale.edu (J.A.S.); C.Best@westernsydney.edu.au (C.T.B.)
5 Department of Linguistics, Yale University, New Haven, CT 06520, USA
6 Haskins Laboratories, New Haven, CT 06520, USA
* Correspondence: jaydene@mail.fresnostate.edu

Abstract: This study tests whether Australian English (AusE) and European Spanish (ES) listeners differ in their categorisation and discrimination of Brazilian Portuguese (BP) vowels. In particular, we investigate two theoretically relevant measures of vowel category overlap (acoustic vs. perceptual categorisation) as predictors of non-native discrimination difficulty. We also investigate whether the individual listener's own native vowel productions predict non-native vowel perception better than group averages. The results showed comparable performance for AusE and ES participants in their perception of the BP vowels. In particular, discrimination patterns were largely dependent on contrast-specific learning scenarios, which were similar across AusE and ES. We also found that acoustic similarity between individuals' own native productions and the BP stimuli were largely consistent with the participants' patterns of non-native categorisation. Furthermore, the results indicated that both acoustic and perceptual overlap successfully predict discrimination performance. However, accuracy in discrimination was better explained by perceptual similarity for ES listeners and by acoustic similarity for AusE listeners. Interestingly, we also found that for ES listeners, the group averages explained discrimination accuracy better than predictions based on individual production data, but that the AusE group showed no difference.

Keywords: acoustic similarity; perceptual similarity; non-native discrimination; non-native categorisation

Citation: Elvin, Jaydene, Daniel Williams, Jason A. Shaw, Catherine T. Best, and Paola Escudero. 2021. The Role of Acoustic Similarity and Non-Native Categorisation in Predicting Non-Native Discrimination: Brazilian Portuguese Vowels by English vs. Spanish Listeners. *Languages* 6: 44. https://doi.org/10.3390/languages6010044

Received: 28 October 2020
Accepted: 18 February 2021
Published: 5 March 2021

1. Introduction

It is well known that learning to perceive and produce the sounds of a new language can be a difficult task for many second language (L2) learners. Models of speech perception such as Flege's Speech Learning Model (SLM; Flege 1995), Best's Perceptual Assimilation Model (PAM, Best 1994, 1995), its extension to L2 acquisition PAM-L2 (Best and Tyler 2007) and the Second Language Linguistic Perception model (L2LP; Escudero 2005, 2009; van Leussen and Escudero 2015; Elvin and Escudero 2019; Yazawa et al. 2020) claim that both the phonological and articulatory-phonetic (PAM, PAM-L2), or acoustic-phonetic similarity (SLM, L2LP) between the native and target language are predictive of L2 discrimination patterns. This suggests that discrimination difficulties are not uniform across groups of L2 learners, at least at the initial stage of learning, as a result of their differing native (L1) phonemic inventories.

When non-native sounds are categorised according to native categories, this is known as a "learning scenario" in the L2LP theoretical framework, as "perceptual assimilation patterns" in PAM, and as "equivalence classification" in SLM. However, it is important to note that whereas L2LP and PAM explore these learning scenarios or assimilation patterns by investigating L2 or non-native phonemic contrasts, SLM focuses on the similarity or dissimilarity between individual L1 and L2 sound categories, rather than contrasts. Specifically, L2LP and PAM posit that contrasts which are present in the native language inventory may be easier to discriminate in the L2 than contrasts which are not present in the L1. This has been demonstrated in Spanish listeners' difficulty to perceive and produce the English /i/–/ɪ/ contrast (Escudero and Boersma 2004; Escudero 2001, 2005; Flege et al. 1997; Morrison 2009). This may be attributed to the fact that the Spanish vowel inventory does not contain /ɪ/ and Spanish listeners often perceive both sounds in the English contrast as one native sound category. The Spanish listeners' difficulty with the English /i/–/ɪ/ contrast can be considered an example of the NEW scenario in L2LP and single-category assimilation in PAM. That is, the two sounds in the non-native contrast are perceived as one single native category. Both models predict that this type of learning scenario (or assimilation pattern) will result in difficulties for listeners when discriminating these speech sounds. Specifically, according to the L2LP framework, the NEW learning scenario is predicted to be difficult because in order for listeners to acquire both sounds (the learning task), a learner must either create a new L2 category or split an existing L1 category (van Leussen and Escudero 2015; Elvin and Escudero 2019).

In contrast, German learners, who have a /i/–/ɪ/ contrast in their L1 vowel inventory, have fewer difficulties when perceiving the contrast in English than Spanish learners (Bohn and Flege 1990; Flege et al. 1997; Iverson and Evans 2007). It is likely that this is an example of the SIMILAR learning scenario in L2LP[1], and PAM's two-category assimilation, whereby the two non-native sounds in the contrast are mapped onto two separate native vowel categories. Both PAM and L2LP would predict that a scenario (or assimilation pattern) of this type would be less problematic for listeners to discriminate than a NEW scenario (or Single Category assimilation) as they can rely on their existing L1 categories to perceive the difference between the L2 phones. In the L2LP framework, the learning task is considered to be easier because learners simply need to replicate and adjust their L1 categories so that their boundaries match those of the L2 contrast (van Leussen and Escudero 2015; Elvin and Escudero 2019; Yazawa et al. 2020). A third scenario, known as the SUBSET scenario (i.e., multiple category assimilation) in L2LP, occurs when one or both sounds in the L2 contrast are perceived as two or more native L1 categories. This scenario may be comparable to focalised, clustered or dispersed uncategorised assimilation in PAM (Faris et al. 2016). While some studies suggest that this learning scenario is not problematic for L2 learners (e.g., Gordon 2008; Morrison 2009, 2003), other studies have shown the SUBSET scenario (or PAM's uncategorised assimilation) to lead to difficulties in discrimination (Escudero and Boersma 2002), particularly when a perceptual or acoustic overlap between the two non-native sounds in the contrast and the perceived native categories occurs (Bohn et al. 2011; Elvin et al. 2014; Tyler et al. 2014; Vasiliev 2013). That is, the two vowels in the contrast are perceived as (or are acoustically similar to) the same multiple categories. For example, Elvin (2016) found that the Brazilian Portuguese vowels /i/ and /e/ were both acoustically similar to and perceived as the same multiple categories in Australian English, namely /iː/, /ɪ/, and /ɪə/. In the L2LP framework, this poses a difficult challenge for learners as they must first realise that certain features or sounds in the target language do not exist, that they cannot process them in the same manner as their L1, and must therefore proceed in a similar manner as with the learning task for the NEW scenario (Elvin and Escudero 2019). The SUBSET scenario can therefore be divided into

1 The L2LP terms "NEW" and "SIMILAR" scenarios differ notably from SLM's use of these terms, and should not be confused with them. The difference in terminology arises from the different foci of the two models: L2LP addresses phonemic contrasts, whereas SLM focuses on individual phones. SLM posits that when listeners are presented with an L2 phone that does not closely resemble any L1 phoneme they form a *new* phonetic category which should be easier to acquire than an L2 phone that is *similar* to an existing L1 phoneme, which should be more difficult to acquire despite the phonetic differences (Colantoni et al. 2015). In contrast, in L2LP a NEW scenario requires the listener to establish a new *contrast* in the L2, which does not exist in the L1, while a SIMILAR scenario reflects a contrast that is similar to one already existing in the L1. SIMILAR scenarios are therefore predicted to be much easier to acquire than NEW scenarios in L2LP.

two categories: SUBSET EASY (or uncategorised non-overlapping in PAM), when the two vowels in the L2 contrast are acoustically similar to and perceived as multiple L1 categories without any overlap, and SUBSET DIFFICULT (or uncategorised overlapping), where the two vowels in the L2 contrast are acoustically similar to and perceives as multiple categories with overlap. A diagram that shows examples of the L2LP scenarios can be seen in Figure 1.

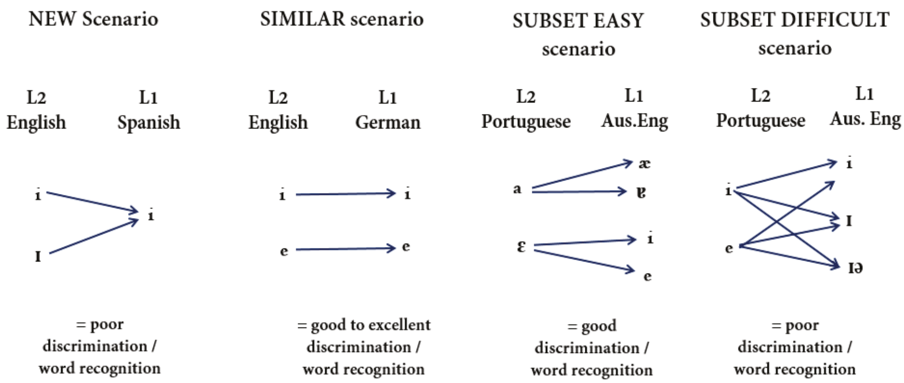


Figure 1. Visual representation of the L2LP learning scenarios.

As the above theoretical models claim, it is the similarity of L2 sounds to native categories that determines L2 discrimination accuracy. It could be the case that individuals whose L1 vowel inventory is larger and more complex than that of the L2 may be faced with relatively less difficulty discriminating L2 vowel contrasts simply because there are many native categories available onto which the L2 vowels can be mapped. Indeed, Iverson and Evans (2007, 2009) found that listeners with a larger vowel inventory (e.g., German and Norwegian) than the L2 were more accurate and had higher levels of improvement post-training at perceiving L2 vowels (e.g., English) than those with a smaller vowel inventory (e.g., Spanish). However, other studies have found that having a larger native vowel inventory than the L2 does not always provide an advantage in L2 discrimination. For instance, recent studies have shown that Australian English listeners do not discriminate Brazilian Portuguese (Elvin et al. 2014) or Dutch (Alispahic et al. 2014; Alispahic et al. 2017) vowels more accurately than Spanish listeners, despite the fact that the Australian English vowel inventory is larger than Brazilian Portuguese and approximately similar in size to that of Dutch, while the Spanish vowel inventory is smaller than those of both Brazilian Portuguese and Dutch. In fact, the findings in Elvin et al. (2014) indicate that Australian English and Spanish listeners found the same Brazilian Portuguese contrasts perceptually easy or difficult to discriminate despite their differing vowel inventory sizes, and that overall, Spanish listeners had higher discrimination accuracy scores than English listeners.

Thus, it seems that vowel inventory size was a good predictor of L2 discrimination performance for some of the aforementioned studies, but not all, which may suggest that this factor alone is not sufficient for predicting L2 discrimination performance. After all, theoretical models such as L2LP and PAM claim that the acoustic-phonetic or articulatory-phonetic similarity between the vowels in the native and target languages, rather than phonemic inventory *per se*, predict L2 discrimination performance. In fact, the L2LP model claims that individuals detect phonetic information in both the L1 and L2 by paying attention to specific acoustic cues (e.g., duration, voice onset time and formants frequencies) in the speech signal. As a result, any acoustic variation in native and target vowel production can influence speech perception (Williams and Escudero 2014). Specifically, the model proposes that the listener's initial perception of the L2 vowels should closely match the acoustic properties of vowels as they are produced in the listener's first language (Escudero and Boersma 2004; Escudero 2005; Escudero et al. 2014; Escudero and Williams 2012). In this way, the L2LP model proposes that both L2 and non-native categorisation patterns and discrimination difficulties can be predicted through a detailed comparison of the acoustic similarity between the sounds of the native and target languages.

This L2LP hypothesis is supported by a number of studies which show that acoustic similarity successfully predicts non-native and L2 categorisation and/or discrimination (e.g., Elvin et al. 2014; Escudero and Chládková 2010; Escudero and Williams 2011; Escudero et al. 2014; Escudero and Vasiliev 2011; Gilichinskaya and Strange 2010; Williams and Escudero 2014). For example, acoustic comparisons successfully predicted that Salento Italian and Peruvian Spanish listeners would categorise Standard Southern British English vowels differently, despite the fact that their vowel inventories contain vowels that are typically represented with the same IPA symbol. The difference was predicted because, despite those shared transcriptions, the acoustic realisations of the five vowels are not identical across the two languages (Escudero et al. 2014). Furthermore, as previously mentioned, Elvin et al. (2014) investigated Australian English and Iberian Spanish listeners' discrimination accuracy for Brazilian Portuguese vowels and found that a comparison of the type and number of vowels in native and non-native phonemic inventories was not sufficient for predicting L2 discrimination difficulties, and that accurate predictions can be achieved if acoustic similarity is considered. Specifically, the L2LP model posits that for the most accurate predictions, the acoustic data should be collected from the same group of listeners intended for perceptual testing. It is this postulate that differentiates L2LP from both PAM/PAM-L2 and SLM, which is the reason we use L2LP as the framework for the current research.

The Present Study

The present study investigates the non-native categorisation and discrimination of Brazilian Portuguese (BP) vowels by Australian English (AusE) and European Spanish (ES) listeners. Similar to Elvin et al. (2014), these language groups were chosen on the basis of their differing inventory sizes. The AusE vowel inventory contains thirteen monophthongal vowels, namely /iː, ɪ, ɪə[2], e, eː, ɜː, ɐ, ɐː, æ, oː, ɔ, ʊ and ʉː/, and is larger than BP, which has seven oral vowels, /i, e, ɛ, a, o, ɔ, u/, while ES has the smallest vowel inventory of the three languages, containing five vowels, /i, e, a, o, u/. Unlike Spanish and Portuguese vowels which are relatively stable in their production, AusE vowels are known to be more dynamic and this has been shown to affect discrimination of some AusE contrasts (see Williams et al. 2018 and Escudero et al. 2018). In this study, we use the L2LP theoretical framework to investigate (1) whether detailed acoustic comparisons using the AusE and ES participants' own native production data successfully predict their non-native categorisation of BP vowels, (2) whether the L2LP learning scenarios identified in non-native categorisation subsequently predict their BP discrimination patterns, and (3) whether measures of acoustic and perceptual (categorisation) overlap are equally good predictors of discrimination accuracy at both group and individual levels (i.e., using individual overlap scores vs. group averages).

While most empirical research in L2 vowel perception investigates L2 development for groups of learners, the present study investigates non-native perception from a group versus an individual perspective. Studies typically focus on learner groups rather than individuals because speech communities have shared linguistic knowledge that allows them to understand each other. As a result, most researchers are particularly interested in how populations behave and how their shared L1 knowledge is relevant to L2 learning. Despite the fact that many researchers are aware that some variability does exist among individuals (Mayr and Escudero 2010; Smith and Hayes-Harb 2011), the group data obtained are generally sufficient for their purposes of demonstrating that shared knowledge

2 The /ɪə/ vowel is traditionally considered a diphthong in Australian English. However, recent studies have shown that this vowel is produced as a monophthong when presented in a closed CVC context (see Elvin et al. 2016) as in this study.

of the sound patterns of the L1 influences L2 speech perception. Importantly, however, other studies (e.g., Díaz et al. 2012; Smith and Hayes-Harb 2011; Wanrooij et al. 2013) have shown that an investigation of individual differences can be important for understanding L2 development. For example, Smith and Hayes-Harb (2011) warn that researchers need to be careful in drawing general conclusions about typical performance patterns for L2 listeners based on group averages, as individual data may be crucial to interpreting group results, especially given the large variety of situations that influence L2 learning by individual learners.

Most of the studies that investigate individual differences in L2 speech perception focus predominately on factors such as age of acquisition, length of residence, language use or motivation (Escudero and Boersma 2004; Flege et al. 1995). In particular, much of the research conducted under the SLM theoretical framework (e.g., Flege et al. 1997, 1995) investigated the above extra-linguistic factors as a means of explaining the degree of foreign accent in an L2 learner. However, even when these factors are controlled, individual differences still seem to persist (Jin et al. 2014; Sebastián-Gallés and Díaz 2012). Furthermore, studies have shown that there are differences in how people hear phonetic cues despite having similar productions that may be related to their auditory processing or their auditory memory (see Wanrooij et al. 2013; Antoniou and Wong 2015). The fact that individual differences persist even when possible factors that influence such variations are controlled suggests that there are real cognitive differences amongst individuals, such as processing style, that influence second language learning. Therefore, language learners, even those at the initial stage (i.e., the onset of learning), may follow different developmental paths to successful acquisition of L2 speech based on their differing cognitive styles and exposure (for a review on recent literature relating to individual differences in processing, see Yu and Zellou 2019). While SLM investigates differences among L2 learners at the level of experiential factors such as age of acquisition and language exposure, the approach is to group learners according to these factors prior to comparing their performance (Colantoni et al. 2015). Studies investigating perception under the framework of PAM also acknowledge the existence of individual differences among listeners; however, few studies are yet to explain such differences. In fact, Tyler et al. (2014) found individual differences in assimilation of non-native vowel contrasts, and proposed that individual variation should be considered when predicting L2 difficulties, but did not examine the sources of the individual differences they had observed. This is where the L2LP model may be particularly relevant: it was specifically designed to account for individual variation among non-native speakers at all stages of learning and across different learning abilities (i.e., perception, word recognition and production). As a result, L2LP predictions can be made for individual learners based on detailed acoustic comparisons of their L1 categories and the categories of the specific target language variety (Colantoni et al. 2015, p. 44).

In our investigation of individual variation, we focus specifically on the fact that individuals from the same native language background may have different acoustic realisations of vowels and this factor may predict individual differences in perceptual performance. That is, the within-category variation in native production may influence non-native categorisation and discrimination. Very few studies (e.g., Levy and Law 2010) have collected vowel productions from the same listeners that they tested in perception, which, according to the L2LP model, is an essential ingredient for accurate predictions of L2 difficulty and for the identification of any individual variation that may be caused by individuals' different acoustic realisations of their own native vowels. Thus, although representative acoustic measurements from the listener populations have successfully explained L2 perceptual difficulty, such comparisons may not account for individual variation among listeners.

The present study reports native acoustic production as well as non-native categorisation and discrimination data from the same participants across all tasks. Although we look at individual versus group data in this study, it is important to note that unlike most other studies of perception and production, the group data we use for perception and production are from the same individuals, which may make the group data more reliable than data for perception and production taken from different groups. Furthermore, the BP acoustic data that we use to measure acoustic similarity are the same recordings that we use as stimuli in the non-native categorisation and discrimination tasks. By doing so, we are able to make predictions relating to the actual stimuli that the participants were presented with, rather than averages taken from other speakers and for vowels in other phonetic contexts. We also control for variation within languages and speakers by ensuring that the participants in each BP-naïve listener group, as well as the speakers in our target BP dialect, were all of similar ages selected from a single urban area within each of their respective countries. By controlling for variation relating to language experience, age and native background, we are able to conduct a carefully controlled investigation of individual differences in non-native perception that may be explained by individual differences in L1 production.

We chose the /fVfe/ context as our target BP stimuli to ensure that our data were comparable to previous studies, specifically Elvin et al. (2014) and Vasiliev (2013). Vasiliev (2013) originally selected target vowels extracted from a voiceless fricative rather than stop context because the voiceless stops differ in VOT (voice onset time) and formant transitions among Spanish, Portuguese, and English. In Elvin et al. (2014), the Australian English acoustic predictions were based on the Cox (2006) corpus, which contained acoustic measurements of adolescent speakers from the Northern Beaches (north of Sydney in New South Wales), collected in the 1990s and extracted from an /hVd/ context. However, Elvin et al. (2016) found that vowel duration and formant trajectories varied depending on the consonantal context in which they were produced. Specifically, vowels produced in the /hVd/ context were acoustically the least similar to the vowels produced in all of the remaining consonantal consonants. Thus, /hVd/ may not be the most representative phonetic context for predicting L2 vowel perception difficulty; in this study, we instead formulated predictions based on native vowels produced in the same phonetic context used as stimuli in testing.

To measure acoustic similarity between vowels, Elvin et al. (2014) used Euclidean Distances between the reported F1 and F2 averages for each vowel. However, because native production data were available for the present study we instead used cross-language discriminant analyses as a method of measuring acoustic similarity, to use in predicting performance in the non-native categorisation and discrimination tasks. This should improve predictions of acoustic similarity over those from simple Euclidean Distance, as we are able to include more detailed acoustic information relevant for vowel perception as input parameters for each individual participant[3].

Considering that patterns of non-native categorisation underlie discrimination difficulties, which according to the L2LP model is predictable based on acoustic properties, the inclusion of non-native categorisation data in the present study further allows for an investigation of whether or not listeners' individual categorisation patterns do in fact predict difficulty in discrimination. The incorporation of a categorisation task also allows us to investigate whether the L2LP learning scenarios at the onset of learning (unfamiliar BP stimuli) are similar across the two listener groups of differing vowel inventories (ES and AusE).

It was essential that we replicated and extended the discrimination task reported in Elvin et al. (2014) with this new set of participants who also completed the native production and non-native categorisation tasks, in order to adequately test the individual difference assumptions of the L2LP model. The L2LP model explicitly states that different listeners have different developmental patterns and it is important to conduct all tasks on the same set of listeners. To this end, we selected naïve listeners in both non-native groups who represent the initial stage of language learning in the L2LP framework. Their inclusion

[3] We note, however, that there are reasons as to why a listener's own productions might not be the best predictors of how they perceive other speakers. Part of a listener's knowledge about vowels includes the ways that different members of their speech community vary (e.g., vocal tract anatomy and social factors). However, we do believe a good way to find symmetry between perception and production is to compare those in the same group of people as in Chládková and Escudero (2012).

provides a good opportunity for assessing differences in language learning ability that are not confounded by other factors that vary widely among actual L2 learners.

The discrimination task in the present study further differs from that reported in Elvin et al. (2014) in that the vowels are presented in a nonce word context rather than as vowels in isolation. We made this change because, outside of the laboratory, learners are faced with words rather than vowels in isolation. The L2LP model assumes continuity between lexical and perceptual development, specifically positing that perceptual learning is triggered when learners attempt to improve recognition by updating their lexical representations (van Leussen and Escudero 2015). Furthermore, if listeners do not interpret the stimuli as speech, which could potentially occur with isolated vowels (particularly synthesized rather than natural vowels), then language-specific L1 knowledge may play less of a role in their perception. That is, listeners from different L1 backgrounds may perceive non-speech in a similar manner but differ in how they perceive the vowels that they perceive to be speech. Given the fact that there were very few group differences in Elvin et al. (2014), it might be that the stimuli were not engaging native language phonology sufficiently reliably for all listeners. Thus, the presentation of vowels in the context of a nonce word not only reflects learning that is closer to a real world situation but also these more speech-like materials allow us to determine whether language-specific knowledge played less of a role in their discrimination of BP.

The present study is therefore, to our knowledge, one of the first to evaluate predictions about L2 perception (both non-native categorisation and discrimination) based on the listeners' own native productions, thereby providing a novel test of one of L2LP's core assumptions. In Section 2, native AusE and ES listeners' native vowel productions are compared to the BP production data that are used as stimuli in the non-native categorisation task (Section 3) and the XAB discrimination task (Section 4). Results from the cross-language acoustic comparisons are used to predict the non-native categorisation patterns in Section 3 and the discrimination results in Section 4. As mentioned above, the participants in the cross-language acoustic comparisons were the same as the participants in the non-native categorisation and discrimination tasks. We do note that the results presented in the cross-language acoustic comparisons and the non-native categorisation tasks are descriptive as we use their categorisation patterns to predict discrimination results in Section 4. In regards to a power analysis of the sample size, for experiment designs with repeated measures analysed with mixed-effects models, Brysbaert and Stevens (2018) recommend a sample size of at least 1600 observations per condition. In our non-native discrimination task, each of the 40 participants completed 40 trials per BP contrast, therefore, this recommendation was met (40 participants × 40 trials = 1600 observations per BP contrast). We do acknowledge a loss of five participants in the non-native categorisation task and we address how this affects our power in our modelling analyses in Section 4.

2. Cross-Language Acoustic Comparisons

2.1. Participants

Twenty Australian English (AusE) monolingual listeners from Western Sydney and twenty European Spanish (ES) monolingual participants from Madrid participated in this study. All participants were Australian English or European Spanish listeners currently residing in Greater Western Sydney or Madrid, respectively, and aged between 18 and 30 years old. The AusE participants reported little to no knowledge of any foreign language. The ES participants reported little to intermediate knowledge of English and little to no knowledge of any other foreign language. AusE participants were recruited through the Western Sydney University psychology pool or from the Greater Western Sydney region, and received $40 AUD for their participation. ES participants were recruited from universities and institutes around the Universidad Nacional de Educación a Distancia and received €30 for their time. All participants were part of a larger-scale study that looked at the interrelations among non-native speech perception, spoken word recognition and non-native speech production. All participants provided informed consent in accordance with the ethical protocols in place at the Universidad Nacional de Educación a Distancia and the Western Sydney University Human Research Ethics Committee.

2.2. Stimuli and Procedure

AusE and ES participants completed a native production task in which they read pseudo-words containing one of the 13 Australian English monophthongs, namely, /iː, ɪ, ɪə, e, eː, ɜː, ɐ, ɐː, æ, oː, ɔ, ʉand ʉː/, or one of five European Spanish vowels, /i, e, a, o, u/, in the /fVf/ (AusE) or /fVfo/ (ES) context. There were 10 repetitions of each vowel, presented in a randomised order, which provided a total of 130 tokens for AusE and 50 tokens for ES per participant. The tokens we used for the analysis of BP vowels were the same as those we used as stimuli in the non-native categorisation and non-native discrimination task. That is they were tokens presented in pseudo-words in the /fVfe/[4] context, produced by five male and five female speakers from São Paulo, selected from the Escudero et al. (2009) corpus. There were a total of 70 BP vowel tokens (one repetition per vowel, per speaker). These BP pseudo-words were produced in isolation and within a carrier sentence e.g., "Fêfe. Em fêfe e fêfo temos ê" which translates to: "Fêfe. In fêfe and fêfo we have ê" Escudero et al. (2009). In our analyses, we selected the vowel in the first syllable of the isolated word which was always stressed and corresponded to one of the seven Portuguese vowels /i, e, ɛ, a, o, ɔ, u/. We used WebMaus (Kisler et al. 2012), an online tool used for automatically segmenting and labelling speech sounds, to segment vowels within each target word in each language (AusE, ES and BP). The automatically generated start and end boundaries were checked and manually adjusted to ensure that they corresponded to the onset/offset of voicing and vocalic formant structure. Vowel duration was measured as the time (ms) between these start and end boundaries. Formant measurements for each vowel token were extracted at three time points (25%, 50%, 75%) following the optimal ceiling method reported in Escudero et al. (2009), in order to ensure that our methods of formant extraction are comparable across both the target and native languages. In the optimal ceiling method, the "ceiling" for formant measurements is selected by vowel and by speaker to minimize variation for the first and second formant values. Formant ceilings ranged between 4500 and 6500 Hz for females and between 4000 and 6000 for males.

2.3. Results: Cross-Language Acoustic Comparisons

Figure 2 shows the average (of all speakers) midpoint F1 and F2 normalised values of the thirteen AusE (black) and five ES (blue) vowels, together with the average (of all speakers) midpoint F1 and F2 normalised values for the BP (purple, circled) vowels that were selected from Escudero et al. (2009) and used as stimuli in the present study. The Lobanov (1971) method was implemented to normalise vowels using the NORM suite (Thomas and Kendall 2007) in R. This specific normalisation method was chosen because it resulted in the best classification performance for the same Brazilian Portuguese vowels used in this study as shown by Escudero and Bion (2007).

Visual inspection of the plot reveals that although AusE has many more vowels in its native vowel inventory than ES, the vowels of both languages fall in and around similar locations along a rough inverted triangle within the acoustic space. Following Strange et al. (2004) and Escudero and Vasiliev (2011), we conducted a series of discriminant analyses as a quantitative measurement of acoustic similarity and used these analyses to predict listeners' non-native categorisation patterns. Before comparing our target language's acoustic similarity with Brazilian Portuguese, we first needed to determine how a trained AusE or ES discriminant analysis model would classify tokens from the same native

[4] Given the fact that the CVCV context is the most common word structure in Spanish, we specifically chose to analyse the /fVfo/ context in the ES native production task. This also prevented the ES participants from producing the target BP stimuli and thereby having an unfair advantage over the AusE participants. We do acknowledge, however, that the post-stressed /e/ in the second syllable of the BP target items is different to the post-stressed /o/ in the second syllable of the Spanish targets, which may have a minor impact on the stressed V acoustic parameters in BP as compared to ES.

[5] See Elvin et al. (2016); Williams et al. (2018) and Escudero et al. (2018) for an overview and visualization of the AusE formant trajectories.

language (known as a cross-validation method). To this end, we fit four separate linear discriminant analysis models: AusE females; AusE males; ES females; ES males. These analyses were conducted to determine the underlying acoustic parameters that predict the vowel categories for test tokens from the BP corpus. The input parameters were F1 and F2 (normalised) values measured at the vowel midpoint (i.e., 50%) as well as duration. We also ran discriminant analyses using F1, F2 and F3 (Bark, duration and formant trajectory as input parameters. We report the results for the discriminant analyses using normalised values as they were more accurate than the values in Bark for both languages. The ES model yielded 98% correct classifications for both males and females, and the AusE model yielded 91.2% (females) and 90.4% (males) correct classifications.

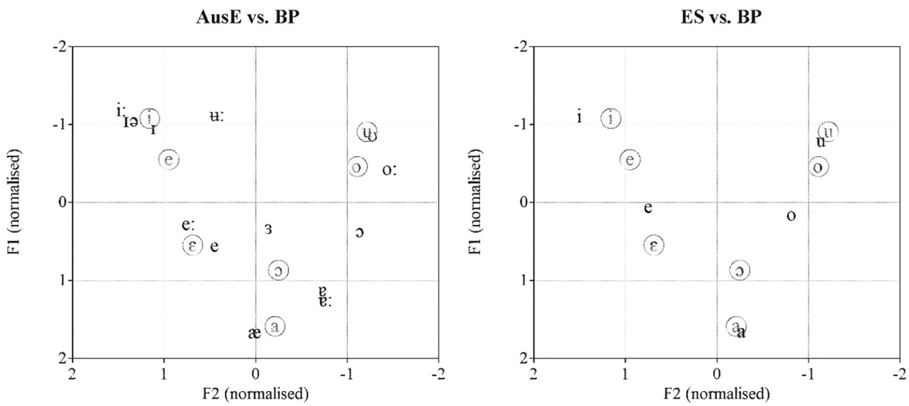


Figure 2. The left panel shows the averaged normalised F1 and F2 values (Hz) for the thirteen AusE[5] (black), and seven BP (gray, circled) vowels. The right panel shows the averaged normalised F1 and F2 values (Hz) for the five ES (black) and seven BP (gray, circled) vowels.

We then conducted a cross-language discriminant analysis, using F1 and F2 normalised values (measured at 50%) and duration as input parameters to determine how likely the BP vowel tokens would be categorised in terms of AusE and ES vowel categories. We fit one model for each individual AusE and ES listener for a total of 40 LDA models. For each individual model, the training data consisted of the 6 tokens of each AusE vowel produced by that same speaker that for which the model was being tested, resulting in a total of 78 native tokens. The test tokens were the same for each of the individual LDA models, that is, 70 male and female BP tokens which were also used as stimuli in the non-native categorisation and discrimination tasks.

In some previous work that has used a typical discriminant analysis, the vowels in the test corpus (in our case BP) are categorised with respect to linear combinations of acoustic variables established by the input corpus (Strange et al. 2004). In other words, the discriminant analysis tests how well the BP tokens can be classified into the vowel categories of the (AusE or ES) input corpus, providing a predicted probability that each vowel token will be categorised as one of the native vowel categories (Strange et al. 2004). Further, the discriminant analysis tests how well the BP tokens fit with centres of gravity of the input corpus tokens (AusE or ES), providing a predicted probability that each vowel will be categorised as one of the native vowel categories. The native vowel category that receives the highest probability for a given BP vowel indicates the native vowel that is acoustically closest to the non-native vowel.

Given the fact that we only have one token per vowel per BP speaker (5 male and 5 female), rather than reporting the overall percentage of times a BP vowel was categorised as a native vowel as is commonly reported (and is usually based on many more tokens), we instead report the probabilities of group membership averaged across the BP vowel tokens: For each individual BP vowel token, we report the predicted probability of it being categorised as any of the 13 native AusE or 5 native ES vowels and average these probabilities over all speakers' tokens for that BP vowel. The benefit of reporting average probabilities across tokens in the present study is that it takes into account that some BP tokens may be acoustically close to more than one vowel, which can be masked by categorisation percentages. The predicted probabilities averaged across the BP tokens for an individual listener and then averaged across all listeners in the AusE and ES groups for AusE and ES are shown in Tables 1 and 2, respectively.

Table 1. Average probability scores of predicted group membership for male and female BP tokens tested on each individual AusE listener model. Probabilities are averaged across the individual discriminant analysis for each speaker. The native vowel category with the highest probability appears in a cell in bold, with no shading, probabilities above chance appear in cells shaded dark grey and probabilities below chance (i.e., 0.08) appear in cells shaded light grey.

AusE Vowels	BP Vowels						
	i	e	ɛ	a	ɔ	o	u
iː	0.22	0.13					
ɪ	**0.71**	**0.59**	0.01				
ɪə	0.05	0.11					
e		0.01	**0.63**		0.16		
eː		0.06	0.33		0.06		
ɜː			0.02	0.05	0.06		
ɐ				0.15	**0.32**		
ɐː				0.06	0.06		
æ			0.02	**0.74**	0.08		
ɔ					0.25	0.26	0.04
oː						0.20	0.04
ʊ					0.01	**0.53**	**0.92**
ʉː	0.02	0.10					

Table 2. Average probability scores of predicted ES vowel group membership for BP male and female tokens tested on the ES model. Probabilities are averaged across the individual discriminant analysis for each speaker. The native vowel category with the highest probability appears in a cell in bold, with no shading, probabilities above chance appear in cells shaded dark grey and probabilities below chance (i.e., 0.20) appear in cells shaded light grey.

ES Vowels	BP Vowels						
	i	e	ɛ	a	o	ɔ	u
i	**0.89**	0.23				0.01	
e	0.11	**0.77**	**1.00**			0.27	
a				**0.99**		0.07	
o				0.01	0.22	**0.64**	0.02
u					**0.78**	0.01	**0.98**

We take the across-individual average probability of vowel group membership to correspond to the degree of acoustic similarity, i.e., a high probability indicates a high level of acoustic similarity. Table 1 shows the averaged probabilities of predicted group membership, which we interpret as being representative of the "average listener". The table of averaged probability scores reveals that each BP vowel showed a strong similarity to a single AusE vowel. However, six of these also showed lower levels of above-chance similarity to one or more other AusE vowels. In the cases where a BP vowel is acoustically similar to two or more AusE vowels, the similarity is not equal across the vowel categories, with the probability scores indicating a greater likelihood of classification of one vowel over the other. An acoustic categorisation overlap can be observed for BP contrasts /i/–/e/ and /o/–/u/, where each vowel in the BP contrast is acoustically similar to the same native AusE vowel(s). In the case of BP /i/–/e/, there is a 0.71 probability that BP /i/ will be categorised as AusE /ɪ/ and a 0.59 probability that BP /e/ will also be categorised as AusE /ɪ/. There is also a 0.22 probability that BP /i/ will be categorised as AusE /iː/, and a probability of 0.13 that BP /e/ will be categorised as AusE /iː/. There is also a 0.11 probability that BP /e/ will be categorised as AusE /ɪə/ as well as a 0.10 probability that it could be categorised as AusE /ʉː/. For BP /o/–/u/ there is a 0.53 probability that BP /o/ will be categorised as AusE /ʉ/ and a 0.92 probability that BP /u/ will be categorised as AusE /ʉ/. We also observe a 0.20 probability that BP /o/ will also be categorised as AusE /oː/. Partial acoustic overlapping is also observed in the BP /o/–/ɔ/ contrast with a 0.26 probability for BP /o/ and a 0.25 probability for BP /ɔ/ to be categorised as AusE /ɔ/. There was also a 0.32 probability that BP /ɔ/ will be categorised as AusE /ɐ/ and we therefore observe a very minimal acoustic overlap with BP /a/. Although there is a 0.74 probability that BP /a/ will be categorised as AusE /æ/, we do see a 0.15 probability that BP /a/ will be categorised as AusE /ɐ/. Finally, we do not see any acoustic overlapping in the BP /a/–/ɛ/.

Table 2 shows the BP tokens tested on the ES model. The results indicate that BP /i/, /ɛ/, /a/ and /u/ are each acoustically similar to different single native categories, namely ES /i/, /e/, /a/, and /u/. while the remaining three BP vowel categories (/e/, /o/ and /ɔ/) show moderate (but much lower) acoustic similarity to a second ES vowel. Similar to the AusE categorisation above, when a BP vowel is acoustically similar to two ES vowels, the similarity is not equal across both categories, with the probability scores indicating a greater likelihood of classification of one vowel over the other. For example, there is a 0.77 probability that BP /e/ would be categorised as ES /e/ and a 0.23 probability of it being categorised as ES /i/. There is also a 0.78 probability that BP /o/ will be categorised as ES /u/ and only a 0.22 probability that it will be categorised as ES /o/. In the case of BP /ɔ/, there is a 0.64 probability that it will be categorised as ES /o/, a 0.27 probability that it would be categorised as ES /e/.

Cases of acoustic overlap can be identified in the ES predicted probabilities of categorisation for the BP contrasts /i/–/e/, /e/–/ɛ/, /o/–/u/ and /o/–/ɔ/. Specifically, in the case of BP /e/–/ɛ/, both vowels were acoustically closest to ES /e/, and in the case of BP /o/–/u/ both vowels were closest to ES /u/. Furthermore, there is a small amount of acoustic overlap in the BP contrasts /i/–/e/ and /o/–/ɔ/. In BP /i/–/e/, while the majority of /i/ tokens and the majority of /e/ tokens were acoustically similar to ES /i/ and /e/, respectively, a smaller percentage of the BP /i/ tokens were acoustically similar to ES /e/ and a smaller percentage of BP /e/ tokens were acoustically similar to ES /i/. A similar result is found with /o/–/ɔ/ as the majority of the BP /ɔ/ tokens were acoustically similar to ES /o/ and a smaller percentage of BP /o/ tokens were acoustically similar to ES /o/. Finally, we do not see any evidence of acoustic overlap for BP /a/–/ɛ/ and /a/–/ɔ/.

2.4. L2LP Predictions for Non-Native Categorisation

The acoustic similarity as determined by the probability scores from our discriminant analyses are used to predict perceived phonetic similarity in a categorisation task. For AusE, there are several cases where the two vowels in the BP contrast are acoustically similar to more than two native categories (in other words, there were predicted probabilities that the BP vowel tokens could be categorised as more than two native categories, with a predicted probability greater than chance). We therefore predict several cases of L2LP's SUBSET EASY and SUBSET DIFFICULT scenarios. Based on the acoustic results averaged across participants, it is likely that all BP vowels will be categorised as more than one native category. For BP /a/–/ɛ/, there is no acoustic overlapping and thus, in non-native categorisation, we expect to find the SUBSET EASY scenario where each BP vowel in the contrast is perceived as more than one native category, but there is no overlapping between the response choices for the two vowels. Where acoustic overlapping occurs, we expect to find the SUBSET DIFFICULT learning scenarios in the non-native categorisation patterns. Specifically, we expect to find perceptual overlap in the non-native categorisation of the BP contrasts /i/–/e/ and /o/–/u/. Partial acoustic overlapping might also lead to instances of the SUBSET DIFFICULT scenario for BP /a/–/ɔ/ and /o/–/ɔ/.

For the ES listeners, we predict on the basis of the LDA results that most BP vowels should be categorised as one single native category. In particular, we expect that BP /i/, and /a/ will be categorised as /i/ and /a/, respectively. Furthermore, we expect to see cases of L2LP's NEW and SIMILAR scenarios. Specifically, we expect to observe instances of the SIMILAR scenario for ES participants in the BP /i/–/e/ and /a/–/ɛ/ contrasts because both vowels in the BP contrast are acoustically similar to separate native categories, with predicted probabilities above 75%. Despite the fact that categorisation of BP /ɔ/ is spread across multiple response categories, we would still predict a case of the SIMILAR scenario for BP /a/–/ɔ/ given the fact that there is no acoustic overlap in the response categories. In contrast, examples of the NEW scenario are predicted for BP /e/–/ɛ/ and /o/–/u/ because both BP /e/ and /ɛ/ are acoustically similar to ES /e/, and both /o/ and /u/ are acoustically similar to ES /u/, with predicted probabilities above 75% It is likely that BP /o/ will be categorised to two native categories as there is a 0.78 probability that it will be categorised as ES /u/ and a 0.22 chance it will be categorised as ES /o/. Finally, BP /ɔ/ should predominately be categorised as ES /o/, but it might also be categorised as ES /e/.

2.5. L2LP Predictions for Non-Native Discrimination

The L2LP model claims that discrimination difficulty can be predicted by the acoustic similarity between native and target language vowel categories, unlike PAM which makes predictions based on articulatory-phonetic similarity and collects perceptual assimilation data and category-goodness ratings to test its predictions. Perhaps the reason that acoustic similarity can be used to predict discrimination difficulty is because acoustic properties and articulation relate to one another (Noiray et al. 2014; Blackwood Ximenes et al. 2017; Whalen et al. 2018). For example, Noiray et al. (2014) have shown that variation in vowel formants correspond closely to variations in the vocal tract area function and even coarser grained articulatory measures such as height of the tongue body. Whalen et al. (2018) compared articulatory and acoustic variability using data from an x-ray microbeam database and found that contrary to popular belief, articulation was not more variable than acoustics, but that variability was consistent across vowels and that articulatory and acoustic variability were related for the vowels. Given this relationship it seems reasonable that acoustic similarity be equal to perceptual similarity in its ability to predict discrimination difficulty. As mentioned in the introduction, we are interested in whether or not acoustic similarity can predict discrimination accuracy and in particular, whether it is comparable to perceptual similarity. One way to measure acoustic and/or perceptual similarity is to calculate the amount of acoustic/perceptual overlap that can be found in a given BP contrast. When two vowels in BP are acoustically/perceptually similar to the same listener vowel category(ies), discrimination of the BP vowels is predicted to be difficult. In this section we use the LDA results to determine how much BP vowels overlap with our listener's native vowel categories and a similar method will be used to measure the amount of perceptual similarity in the non-native categorisation task. We predict that the perceptually easy contrasts for both groups of listeners to discriminate would be those with little to no acoustic overlap (i.e., the two vowels in the BP contrast are acoustically similar to different native categories). The BP contrasts with a large amount of acoustic overlap (i.e., the two BP vowels in the contrast are acoustically similar to the same native category(ies)) should be difficult to discriminate.

To quantify acoustic overlap, we adopted Levy's (2009) "cross language assimilation overlap" method. This method provides a quantitative score of overlap between the members of a non-native contrast and native categories. Although originally designed to compute perceptual overlap scores based on listeners' perceptual assimilation patterns for testing predictions in PAM (which we do in fact apply to our categorisation data), we use our LDA results. Each overlap score was calculated by adding categorisation probabilities in cases where the two vowels in the BP contrast were categorised as the same native categories. This gives an aggregate probability of perceiving the two BP vowels as the same native category. For example, in the case of BP /i/–/e/, as observed in Table 1, there was a non-zero probability that both BP /i/ and BP /e/ would each be categorised as AusE /iː/, /ɪ/, /ɪə/ and /ʉː/. To calculate the overlap score for this contrast, we took the smaller proportion of when both BP vowels had a probability of being categorised as the same AusE vowel category for each native vowel and add those together. Thus, in the case of AusE /ɪ/ there was a 0.71 probability that BP /i/ would be categorised as this vowel and a 0.59 probability that BP /e/ would also be categorised as AusE /ɪ/. The smaller proportion in this case would be 0.59 for BP /e/, as well as AusE /iː/ (0.13), AusE /ɪə/ (0.05) and AusE /ʉː/ (0.02), which were included in the calculation of the acoustic overlap to obtain an acoustic overlap score of 0.79. Thus, summing together each of the smaller proportions, the calculation of acoustic overlap for BP /i/–/e/ was as follows: AusE /ɪ/0.59 + AusE /iː/0.13 + AusE /ɪə/0.05 + AusE /ʉː/0.02 = 0.79 acoustic overlap. Table 3 shows the acoustic overlap scores for each language.

Table 3. Acoustic overlap scores for AusE and ES listeners.

Group	BP Vowel Contrast					
	/a/–/ɔ/	/a/–/ɛ/	/i/–/e/	/o/–/u/	/e/–/ɛ/	/o/–/ɔ/
AusE	0.34	0.04	0.79	0.61	0.08	0.26
ES	0.08	0.00	0.34	0.80	0.77	0.23

Based on the acoustic overlap scores in Table 3, we would predict that the BP contrasts with little to no acoustic overlap would be perceptually easier to discriminate than those with higher overlap scores. In particular, both groups should find BP /o/–/u/ difficult to discriminate and BP /a/–/ɛ/ easy to discriminate. However, these acoustic comparisons do predict group differences: ES listeners should find BP /a/–/ɔ/ and /i/–/e/ easier to discriminate than AusE listeners, whereas BP /e/–/ɛ/ should be easier for AusE than ES listeners to discriminate.

3. Non-Native Categorisation

3.1. Participants

The participants in the non-native categorisation task were the same as those previously reported in the cross-language acoustic comparisons. However, the non-native categorisation results of five ES participants were excluded due to an error that occurred during testing.

3.2. Stimuli and Procedure

Participants were presented with the same BP pseudo-words that served as the test data for the discriminant analyses in the cross-language acoustic comparisons. There were a total of 70 /fVfe/ target words (7 vowels $\times$ 10 speakers), as well as three additional nonsense words by each speaker (/pipe/, /kuke/ and /sase/), included as filler words. Thus, in the non-native categorisation task we had a total of 100 BP word tokens (70 target and 30 fillers).

In keeping with Vasiliev (2013; see also, e.g., Tyler et al. 2014), this vowel categorisation task followed the discrimination task (reported in the next section) because we wanted to avoid any familiarisation with the natural stimuli in the discrimination task. We present the results of the categorisation task first because these results are used to make predictions about discrimination. In the categorisation task, participants categorised the stressed vowel sound of each target BP word (i.e., the target vowel) to one of their own 13 AusE or 5 ES vowel categories. Unlike Spanish, English is not orthographically transparent. Thus, while the ES listeners saw the 5 vowel categories (i, e, a, o, u) on the screen, the AusE vowels were presented in one of the 13 keywords, heed, hid, heared, head, haired, heard, hud, hard, had, hoard, hod, hood and who'd, which correspond to the AusE phonemes /iː, ɪ, ɪə, e, eː, ɜː, ɐ, ɐː, æ, oː, ɔ, ʉ, ʉː/, respectively. Participants heard each target and filler item once, and were required to choose one of their own native response options on each trial, even when unsure. The task did not move on to the next trial until a response had been chosen. All trials were presented in a randomised order. Participants received a short practice session before beginning the task and took approximately 10 min to complete it.

3.3. Results

Tables 4 and 5 present the percentage of times each BP vowel was categorised by each group as a native AusE or a native ES vowel, respectively.

Table 4. Australian English listeners' classification percentages. The native vowel category with the highest classification percentage appears in bold, classification percentages above chance appear in cells shaded dark grey and classification percentages below chance (i.e., 0.08) appear in cells shaded light grey.

AusE Vowels	BP Vowels						
	i	e	ɛ	a	o	ɔ	u
iː	**0.43**	**0.23**	0.05				0.01
ɪ	**0.43**	0.20	0.02				
ɪə	0.06	0.15	0.06	0.02		0.01	0.01
e	0.06	0.14	0.14				
eː	0.01	**0.23**	**0.58**	0.06	0.01	0.04	
ɜː			0.01	0.01	0.08	0.05	0.03
æ	0.01	0.03	0.07	0.38		0.01	
ɐː	0.01	0.02	0.08	**0.50**		0.13	
ɐ				0.03	0.04		0.06
ɔ					0.09	0.15	0.07
oː		0.02	0.01	0.02	**0.53**	**0.58**	0.06
ʊ					0.23	0.03	**0.65**
ʉː					0.04	0.02	0.12

Table 5. European Spanish listeners' classification percentages. The native vowel category with the highest classification percentage appears in bold, classification percentages above chance appear in cells shaded dark grey and classification percentages below chance (i.e., 0.20) appear in cells shaded light grey.

ES Vowels	BP Vowels						
	i	e	ɛ	a	o	ɔ	u
i	**1.0**	**0.56**	0.01				
e		0.44	**0.94**				
a			0.05	**1.00**		0.02	
o					**0.69**	**0.97**	0.01
u					0.31	0.01	**0.99**

The categorisation percentages reported in Table 4 (AusE) and Table 5 (ES) are in line with our prediction based on acoustic similarity that most BP vowels would be categorised into more than two native categories by AusE listeners, and that most BP vowels would instead be categorised into one single native category by ES listeners.

Indeed, as predicted, all BP vowels were categorised as two or more native AusE categories, and there was some evidence of perceptual overlap between the expected pairs of BP vowels. In particular, we found examples of the SUBSET DIFFICULT scenario in the BP contrasts /i/–/e/, /e/–/ɛ/, /o/–/ɔ/ and /o/-u/, where both vowels in the contrast were categorised into two or more of the same native AusE vowels.

With respect to the BP /i/–/e/ contrast, BP /i/ was categorised as AusE /iː/ as well as AusE /ɪ/, 43% of the time for both vowels. This finding was predicted by acoustic similarity, however instead of there being a larger percentage of categorisation to AusE /ɪ/ as expected, categorisation was split equally across the two AusE categories. As for BP /e/, our discriminant analysis indicated it would be categorised across four native vowel categories, namely /iː/, /ɪ/, /ɪə/ and /ʉː/, with the largest classification percentage predicted to be to AusE /ɪ/. This prediction was largely consistent with AusE listeners' non-native categorisation patterns, with /e/ being categorised as AusE /iː/ (23%), /ɪ/ (20%) and /ɪə/ (15%). Although the discriminant analysis prediction was that BP /e/ would be categorised as AusE /ʉː/ to a small extent (10%), this was not observed, and listeners instead rather substantially categorised the vowel to two unpredicted AusE vowels, /e/ (14%, i.e., as often as to /ɪə/) and /eː/ (23%, i.e., equal to the actual choices of the top acoustically predicted choice /iː/).

As for the BP contrast /e/–/ɛ/, BP /ɛ/ tokens were expected to be predominately categorised as AusE /e/ and /eː/, two of the AusE vowels to which BP /e/ was categorised. This was indeed the case as BP /ɛ/ was categorised as AusE to /eː/58% of the time and to /e/ 14% of the time.

For the BP /o/–/ɔ/ contrast, a large majority of BP /o/ tokens were acoustically predicted to be categorised as AusE /ʉ/, with a much lower probability that some would also be categorised as AusE /oː/ and /ɔ/. However, the listeners actually reversed the balance between the two AusE vowel categories: the large majority BP /o/ tokens were instead categorised as AusE /oː/ (53% of the time), and as AusE /ʉ/ only 23% of the time. Furthermore, our acoustic predictions suggested that BP /ɔ/ tokens would be categorised as a number of different AusE vowels, specifically AusE /ɐ/, /ɔ/ and /e/. However, the non-native categorisation patterns indicate that the great majority of BP /ɔ/ tokens were categorised as AusE /oː/ (58% of the time), with only 15% being categorised as /ɔ/ and 13% as /ɐː/, and none selected AusE /e/.

Finally, with respect to the BP /o/-u/ contrast, while our acoustic analysis successfully predicted the majority of BP /u/ tokens to be categorised as AusE /ʉ/ (65%), interestingly 12% of the tokens were categorised /ʉː/, which was not predicted to be a listener choice. Recall that the results of the discriminant analysis indicated a 10% probability that BP /e/ would be categorised as AusE /ʉː/ but this did not occur, yet conversely, here we see that BP /u/ was categorised as AusE /ʉː/12% of the time, although it was not predicted acoustically.

For the BP /a/–/ɛ/ and /a/–/ɔ/ contrasts, our results are partially consistent with the patterns of acoustic similarity. We found that BP /a/ was categorised as /æ/38% of the time. However, instead of the predicted moderate level of choice of AusE /æ/ (as in had) for BP /a/, the long AusE /ɐː/ (as in heart) was instead selected most frequently at 50% of the time. Given that no perceptual overlap is observed in BP /a/–/ɛ/, the categorisation pattern would correspond to a SUBSET EASY scenario. The same might also apply to BP /a/–/ɔ/. However, we do see a partial overlap, with 13% of the BP /ɔ/ perceived as AusE /ɐː/ (which was the most frequent response for BP /a/).

The minor discrepancies between the acoustic predictions and the categorisation results could be related to the fact that we selected the best fitting discriminant model that in this case did not include F3, which conveys information related to lip rounding. It may be that although in machine learning vowels can be categorised with high accuracy using duration and normalised F1 (height) and F2 (backness) values only, human listeners may not be able to help but pay attention to other aspects of the signal. Thus, human listeners may primarily use F3 when it cues rounding, an articulatory property, but ignore it (or give it unequal weight) in cases where rounding is not present. This therefore suggests that listeners seem to perceive rounding (as opposed to F3), and the likely reason they did not choose /ʉː/ in the categorisation of BP /e/ is that they did not detect the rounding that they may have detected when categorising BP /u/.

Turning to the categorisation results for ES presented in Table 5, we found that acoustic similarity was largely consistent in predicting ES listeners' non-native categorisation patterns. As expected, BP /i/, /ɛ/, /a/ and /u/ were each categorised as the different single native ES vowel categories identified in acoustics, namely /i/, /e/, /a/ and /u/, respectively.

Also in accordance with our acoustic analyses, BP /e/, /o/ and /ɔ/ showed some degree of categorisation to more than one ES category. BP /e/ was categorised as ES /i/ and /e/ as expected. BP /o/ was categorised as both ES /o/ and /u/ as expected. However, the majority of tokens were categorised as ES /o/ instead of ES /u/, reversing the acoustic prediction. Finally, our discriminant analysis predicted a 64% probability that BP /ɔ/ would be categorised as ES /o/ and only a 27% probability that it would be categorised as ES /e/. However, the non-native categorisation task indicated that 97% of the /ɔ/ tokens were categorised by actual listeners as ES /o/. Thus, here we also see a case where our acoustic prediction regarding the categorisation of BP /ɔ/ to ES /e/ was not borne out. Again, it seems as though this discrepancy could be explained by the amount of rounding in the BP /ɔ/ and /o/vowels. Recall that the best fitting LDA for our data was the one that included normalised duration and F1 and F2 values only, thus it did not take into consideration F3, which as mentioned above usually corresponds to lip rounding. Although our acoustic analysis for ES indicated a potential categorisation of BP /ɔ/ to ES /e/, it is not surprising that this was not the case as BP /ɔ/ is a rounded vowel whereas ES /e/ is not. It seems likely that human listeners would weight rounding heavily in their categorisation of BP /ɔ/, i.e., hear it as ES /o/ rather than ES /e/. Furthermore, the results indicate that SIMILAR and NEW L2LP scenarios are represented in these non-native categorisation results. Non-native categorisation of BP /a/–/ɛ/ and /a/–/ɔ/ both show evidence of the SIMILAR scenario, as neither BP contrast yielded perceptual overlapping in the ES response categories. For the remaining contrasts, we see evidence of L2LP's NEW scenario as the perceptual overlapping occurred over just one single response category.

3.4. Discussion

Based on the above findings, it appears that the non-native categorisation patterns for ES listeners were largely in line with predictions based on acoustic similarity between the target BP vowels and the listeners' production of their native vowel categories reported in the cross-language acoustic comparisons.

For AusE listeners, however, there seemed to be more discrepancies between acoustic predictions and categorisation patterns. For example, there were some cases where the acoustic analyses indicated a small probability that AusE /ʉː/ would be a likely response category and it was not, or vice versa. We also observed that BP /a/ tokens were categorised more frequently as AusE /ɐː/ rather than AusE /æ/, which was acoustically predicted to be the most likely response category. These differences between predicted and actual categorisation could be related to cue weighting as well as to dynamic features in AusE vowels. The L2LP theoretical framework includes a strong emphasis on acoustic and auditory cue-weighting (the relative importance of acoustic cues in the learner's native and target languages). Thus, it may be that listeners weight certain cues (e.g., lip rounding or duration) more than others, as has been shown in previous studies (e.g., Curtin et al. 2009).

Studies have also shown that AusE vowels are marked by dynamic formant features (Watson and Harrington 1999; Elvin et al. 2016; Escudero et al. 2018), thus it could be that listeners are searching for these dynamic features in order to categorise the target BP vowels. Future studies may consider improving the acoustic analyses by measuring the amount of spectral change in the native and target language and running discriminant analyses on those data. For example, Escudero et al. (2018) measured the amount of spectral change in three AusE vowels (/i/, /ɪ/ and /uː/) by extracting formant values at 30 equally spaced time points. Discrete cosine transform (DCT) coefficient values were obtained, which correspond to the vowel shape in the F1/F2 space (formant means, magnitude and direction). These DCT values were used in discriminant analyses, resulting in better overall categorisation of the AusE vowels than discriminant analyses run on F1, F2 and F3 values alone. Thus, using DCT values for the native and target language may provide more reliable acoustic predictions that correspond more closely to actual human performance in non-native categorisation.

Predictions for Discrimination Accuracy

In order to predict listeners' performance in discrimination, we calculated perceptual overlap scores based on the amount of overlapping in the listeners' non-native categorisation of the BP vowels. We computed the perceptual overlap scores following Levy (2009) for our categorisation data to determine how predictions of discrimination difficulty based on non-native categorisation would compare with our predictions based on our acoustic comparisons as described in the cross-language acoustic comparisons (see Table 3). To determine a perceptual overlap score based on the categorisation percentages reported in Tables 4 and 5, we sum the smaller percentages of when both BP vowels in a given contrast are categorised as the same native vowel category. Table 6 presents the acoustic and perceptual overlap scores.

Table 6. Acoustic and perceptual overlap scores for AusE and ES listeners expressed as proportions.

Group	Overlap	BP Contrasts					
		/a/–/ɔ/	/a/–/ɛ/	/i/–/e/	/o/–/u/	/e/–/ɛ/	/o/–/ɔ/
AusE	**Acoustic**	0.34	0.04	0.79	0.61	0.08	0.26
	Perceptual	0.22	0.11	0.58	0.47	0.57	0.73
ES	**Acoustic**	0.08	0.00	0.34	0.80	0.77	0.23
	Perceptual	0.02	0.05	0.56	0.32	0.45	0.70

When comparing the predictions for discrimination difficulty based on perceptual overlap scores with those based on acoustic overlap, the predictions are rather similar for BP /a/–/ɛ/. That is, both acoustic similarity and non-native categorisation patterns predict that this contrast should be perceptually easy for both groups of listeners to discriminate. That is because this contrast appears to correspond to the L2LP SUBSET EASY learning scenario for AusE listeners and the SIMILAR learning scenario for Spanish listeners. Acoustic similarity and categorisation patterns indicate the same L2LP scenarios to apply to BP /a/–/ɔ/ in both languages, and so this contrast should also be perceptually easy to discriminate.

For the remaining four contrasts, predictions based on acoustics and non-native perceptual categorisation differ. Acoustic similarity predicts BP /i/–/e/ to be perceptually difficult for AusE listeners to discriminate, due to the L2LP SUBSET DIFFICULT, but the ES categorisation results suggest that this is also likely to be difficult for ES listeners to discriminate as there is evidence of the L2LP NEW scenario. Predictions based on acoustic similarity predict that both groups should find BP /o/–/u/ to be one of the most difficult contrasts to discriminate, whereas perceptual overlap scores suggest that /o/–/ɔ/ should be the most difficult to discriminate. Difficulties for both contrasts is predicted by the L2LP SUBSET DIFFICULT scenario for AusE listeners and the NEW scenario for ES listeners.

From these findings, two possible predicted patterns of difficulty can be identified. Predictions based on acoustic similarity would suggest the following order of difficulty for the two groups (ranging from the lowest acoustic overlap score to the highest):

AusE: /a/-/ɛ/>/a/-/ɔ/>/e/-/ɛ/>/o/-/ɔ/>/i/-/e/~/o/-/u/

(1)ES: /a/-/ɛ/~/a/-/ɔ/>/e/-/ɛ/>/o/-/ɔ/~/i/-/e/~/o/-/u/

On the other hand, non-native categorisation patterns, i.e., perceptual similarities, would predict that both AusE and ES listeners would share the same pattern of difficulty:

AusE and ES: /a/-/ɛ/>/a/-/ɔ/>/o/-/u/>/e/-/ɛ/>/i/-/e/>/o/-/ɔ/

In all cases, BP /a/–/ɛ/ is predicted to be the easiest to discriminate, with the order of difficulty differing among the rest of the contrasts for the acoustic and perceptual predictions. An examination of the pattern of difficulty in the results for discrimination accuracy will shed light on whether discrimination difficulty is in line with predictions based on acoustic similarity or those based on non-native categorisation patterns.

4. Non-Native Discrimination

4.1. Participants

Participants in this task were the same 20 AusE and 20 ES participants previously reported in the cross-language acoustic comparisons and non-native categorisation task[6].

4.2. Stimuli and Procedure

Listeners were presented with the same 70 naturally produced BP /fVfe/ target words (7 vowels × 10 speakers), selected from Escudero et al.'s (2009) corpus, previously reported and analysed in the cross-language acoustic comparisons and the non-native categorisation task.

To test for discrimination accuracy, participants completed an auditory two-alternative forced choice (2AFC) task in the XAB format, similar to that of Escudero and Wanrooij (2010), Escudero and Williams (2012) and Elvin et al. (2014). The task was run on a laptop using the E-Prime 2.0 software program.

Three stimulus items were presented per trial. The second (A) and third (B) items were always from different BP vowel categories and the first item (X) was the target item about which a matching decision was required. In each trial, X was always one of the 70 target BP words, produced by the five male and five female speakers reported above. The A and B stimuli were always the seventh male and seventh female speaker from the Escudero et al. (2009) corpus to avoid any confusion of overlapping target stimuli and response categories. The gender of the A stimuli was always the same gender of the speaker of the B stimuli. This differs from the Elvin et al. (2014) study where the A and B stimuli were synthetic. Furthermore, the order of the A and B responses was counterbalanced (namely, XAB and XBA). On each trial, participants were instructed to listen to the three words using headphones and were required to make a decision as to whether the first word they heard sounded more like the second or the third.

For the first ten participants for each language group, testing consisted of six blocks of categorical discrimination tasks, with a short break permitted between blocks. Each block consisted of 40 trials with one of the six BP contrasts, namely /a/–/ɔ/, /a/–/ɛ/, /i/–/e/, /o/–/u/, /e/–/ɛ/ and /o/–/ɔ/, with the blocks presented in a randomised order. To determine whether discrimination accuracy differs when stimuli are blocked by contrast or randomised, the remaining 10 participants per group completed the same

[6] We note that the five participants missing from the non-native categorisation are included in our analyses of non-native discrimination. We do address the issue of the missing data from the categorisation task in our comparison of cross-linguistic acoustic similarity vs. perceptual similarity below.

discrimination task, with the same breaks, but with the stimulus contrasts presented in random order, unblocked.

4.3. Results

We conducted a repeated-measures Analysis of Variance (ANOVA), with contrast as a within-subjects factor and condition (blocked, randomised) as a between-subjects factor, to evaluate whether blocking by BP contrast has an effect on overall performance. The results yielded no significant effect of condition on performance, [F (1, 38) = 1.905, $p = 0.176$, η $p^2 = 0.048$], suggesting that listeners had similar accuracy scores regardless of the condition (blocked vs. randomised). Thus, Figure 3 shows discrimination accuracy for the AusE and ES groups, including their variability, across the six BP vowel contrasts for both conditions pooled together.

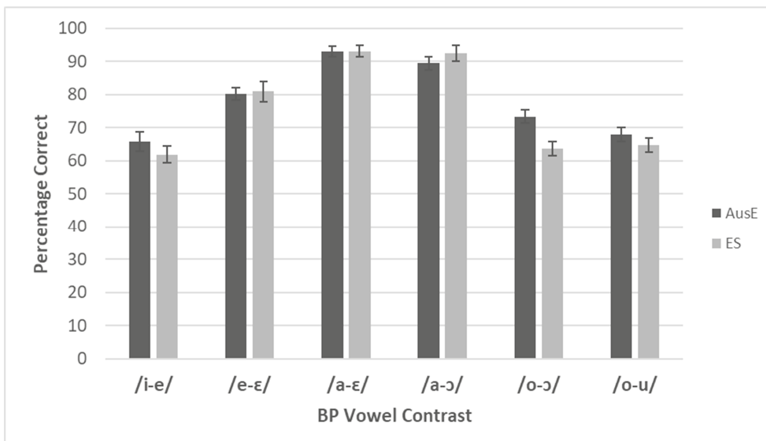


Figure 3. Overall discrimination accuracy including variability for AusE and ES.

The figure shows that the average accuracy scores are comparable across the two language groups. Both groups appear to have highest accuracy for /a/–/ɔ/ and /a/–/ɛ/ and lowest accuracy for /i/–/e/, /o/–/ɔ/ and /o/–/u/, with intermediate accuracy on /e/–/ɛ/.

In order to test for differences across the contrasts and between the two groups, a linear mixed-effects binary logistic model was conducted in SPSS with participant, X stimulus and trial included as random effects and BP contrast and language group included as fixed effects. Recall that for experiment designs with repeated measures analysed with mixed-effects models, Brysbaert and Stevens (2018) recommend a sample size of at least 1600 observations per condition. As each of the 40 participants completed 40 trials per BP contrast, this recommendation was met (40 participants × 40 trials = 1600 observations per BP contrast).

The model revealed a significant main effect of contrast [χ^2 (5, N = 9599) = 646.212, $p \leq 0.001$]. This significance is based on a comparison of nested models by the likelihood ratio test. There was no significant effect for language group [χ^2 (1, N = 9599) = 0.880, $p = 0.348$]. However, the interaction of BP contrast*language group [χ^2 (5, N = 9599) = 19.35, $p = 0.002$] was significant. This confirms that discrimination accuracy varies depending on the BP contrast and that although there are no reliable differences between AusE and ES in terms of overall accuracy, the two groups did differ in their performance on some of the BP contrasts. We ran Fisher's LSD-corrected post-hoc pairwise comparisons to determine the group differences across the contrasts and found that the ES listeners had higher accuracy than AusE participants for discrimination of BP /a/–/ɔ/ ($p = 0.035$, 95% CI [−0.06, 0.00]), whereas the AusE participants performed better than the ES participants on BP /o/–/ɔ/ ($p \leq 001$, 95% CI [0.05, 0.14]).

Fisher's LSD-corrected post-hoc pairwise comparisons were also used to compare discrimination accuracy for each language group across the six BP contrasts. The results indicated that both groups found the same contrasts equally easy/difficult to discriminate. In particular, both groups had significantly higher accuracy scores for BP /a/–/ɛ/ than the remaining contrasts (AusE: all *p*s $\leq$ 0.013, ES: all *p*s $\leq$ 0.001), with the exception of the ES listeners' performance on BP /a/–/ɔ/ which was comparable to BP /a/–/ɛ/ (p = 0.628). The results further indicated that both groups found BP /a/–/ɔ/ to be significantly easier to discriminate than the remaining four contrasts (/i/–/e/, /o/–/u/, /e/–/ɛ/ and /o/–/ɔ/) (AusE and ES: all *p*s $\leq$ 0.001). The AusE participants had significantly lower accuracy scores for BP /i/–/e/ and /o/–/u/ than the other four BP contrasts (all *p*s $\leq$ 0.018), but comparable levels of difficulty among the latter four contrasts (p = 0.339). Likewise, the ES participants had comparable levels of difficulty for BP /i/–/e/ and /o/–/u/, but also /o/–/ɔ/ (*p*s = 0.233–0.676), with significantly lower accuracy scores on these contrasts than the remaining three contrasts (all *p*s $\leq$ 0.001). The results indicate that there was no significant difference between BP /e/–/ɛ/ and BP /a/–/ɔ/or /a/–/ɛ/ for both AusE and ES listeners. However, BP /e/–/ɛ/ was significantly easier for both groups to discriminate than the remaining three contrasts (AusE and ES: all *p*s $\leq$ 0.001). Based on the results from the statistical analyses, the order of difficulty from least difficult to most difficult (where "~" means equal or comparable difficulty and ">" signifies higher accuracy) is as follows:

AusE /a/-/ɛ/>/a/-/ɔ/>/e/-/ɛ/>/o/-/ɔ/>/i/-/e/~/o/-/u/

(2)ES /a/-/ɛ/~/a/-/ɔ/>/e/-/ɛ/>/o/-/ɔ/~/i/-/e/~/o/-/u/

4.4. Acoustic vs. Perceptual Similarity as a Predictor of Non-Native Discrimination

Recall that we predicted two possible patterns of discrimination difficulty, depending on whether or not findings would be more consistent with predictions based on acoustic similarity or those based on non-native categorisation patterns as determined by the degree of perceptual overlap. As predicted by both acoustic similarity and perceptual overlap, BP /a/–/ɛ/ was indeed easiest for both groups to discriminate. We also find that in line with the acoustic and perceptual overlap predictions, the BP /a-ɔ/ contrast was indeed perceptually easy for ES listeners. However, it was also perceptually easy for AusE listeners. As predicted acoustically, BP /i/–/e/ was indeed difficult for AusE listeners and, in line with the perceptual overlap predictions, this contrast was also difficult for ES listeners. We also find that in line with our acoustic predictions, BP /o/–/u/ was difficult for both groups to discriminate. In comparison, the AusE listeners' results for BP /e/–/ɛ/ and /o/–/ɔ/ were more in line with predictions based on the perceptual overlap scores of their non-native categorisation patterns.

To assess quantitatively how different measures of vowel category overlap (acoustic vs. perceptual) relate to discrimination, we fit mixed-effects binomial logistic regression models using the glmer function (binomial family) in R (3.5.1). Accuracy was the dependent variable (correct vs. incorrect) and either perceptual overlap or acoustic overlap was the predictor (fixed factor). Rather than use the raw values for the predictor, acoustic and perceptual overlap scores for each BP contrast were rank coded from least overlap (=1) to greatest overlap (=6) in light of Levy's (2009) treatment of the overlap scores as ordinal and not as interval measures. For any instances of a tie, the average rank was assigned (as shown in Table 7). Subsequently, overlap was centred around the middle of the ranking scale, meaning that the models' intercepts represent average accuracy between ranks 3 and 4 and that the fixed effect of overlap represents the average decrease in accuracy associated with a one-unit increase in overlap rank.

Table 7. Ranking of the 6 BP contrasts according to perceptual and acoustic overlap scores. Individual overlap rankings represent the mean rank across participants. In order to compare the overlap scores to the actual discrimination results, we provide the order of discrimination difficulty (bottom two rows) from the group discrimination results across the six BP contrasts reported from the previous page.

	Perceptual Overlap Rank				Acoustic Overlap Rank			
	Group		Individual		Group		Individual	
BP Contrast	**AusE**	**ES**	**AusE**	**ES**	**AusE**	**ES**	**AusE**	**ES**
/a/–/ɔ/	2.00	1.00	1.95	1.60	4.00	2.00	4.05	2.90
/a/–/ɛ/	1.00	2.00	2.00	1.87	1.00	1.00	1.48	1.17
/i/–/e/	5.00	5.00	4.33	4.40	6.00	4.00	5.40	4.03
/o/–/u/	3.00	3.00	3.88	3.83	5.00	6.00	4.80	4.80
/e/–/ɛ/	4.00	4.00	3.88	4.17	2.00	5.00	2.18	4.30
/o/–/ɔ/	6.00	6.00	4.98	5.13	3.00	3.00	3.10	3.80
	Order of Discrimination Difficulty							
AusE	/a/–/ɛ/>/a/–/ɔ/>/e/–/ɛ/>/o/–/ɔ/>/i/–/e/~/o/–/u/							
ES	/a/–/ɛ/~/a/–/ɔ/>/e/–/ɛ/>/o/–/ɔ/~/i/–/e/~/o/–/u/							

The random factors were Participant (with random slopes for either perceptual or acoustic overlap rank, as this factor was repeated across listeners), Item (the X Stimulus-Contrast combination with random slopes for either perceptual or acoustic overlap rank, as this factor was repeated across items) and Trial.

As five ES listeners lacked perceptual assimilation data, for these participants, we used the mean individual perceptual overlap values from the remaining ES listeners and ranked the six BP contrasts accordingly. For the ES model on individual perceptual overlap, we checked whether controlling for the subgroup of five ES listeners with imputed individual overlap scores would provide a closer model fit. To do so, a likelihood ratio test was conducted comparing a model not controlling for the subgroup and a model including an effect of subgroup (the five ES listeners versus the remaining ES listeners) and its interaction with individual perceptual overlap. This showed that the more complex model provided almost no improvement over the simpler model (χ^2 (2) = 0.42, p = 0.81).

Before accepting the results of the mixed-effects models, we tested whether they were sufficiently powered to detect the smallest meaningful effect size of perceptual or acoustic overlap. This was because the models were run on the two groups' data separately unlike the previous analysis examining discrimination accuracy with both groups together, meaning there were far fewer than the recommended 1600 observations per condition (Brysbaert and Stevens 2018). We defined the smallest meaningful effect size as one fewer correct response with each one-unit increase in overlap rank (equivalent to 2.5% of trials within each listener's set of responses per BP contrast). For each model, using the SIMR package in R (Green and MacLeod 2016), 1000 Monte Carlo simulations were run where correct and incorrect responses were randomly generated such that the regression coefficient for the smallest meaningful effect of overlap rank remained the same. We deemed a model to have sufficient power if at least 80% of its simulations detected this smallest effect with a p-value less than 0.05. All models passed this test.

The results from the mixed models presented in Table 8 indicate that the level of acoustic overlap and perceptual overlap based on both individual and group calculations indeed influenced the participants' discrimination accuracy. This means that these measures can be reliably used to predict discrimination difficulty. To examine whether one measure of overlap (acoustic vs. perceptual and group vs. individual) better explained our discrimination data, we conducted pairwise comparisons on the Bayesian Information Criterion (BIC) from each model. BIC is intended for model selection and takes into account the log-likelihood of a model and its complexity. To quantify the weight of evidence in favour of one model over an alternative model, Bayes Factors (BFs) can be computed based on each model's BIC (Wagenmakers 2007). BFs < 3 provide weak evidence, BFs > 3 indicate positive support and BFs > 150 indicate very strong support for the alternative model (Wagenmakers 2007). For AusE listeners, the models containing group or individual acoustic overlap scores were very strongly supported over their counterpart models containing perceptual overlap scores (BFs > 150). For ES listeners, on the other hand, the opposite was the case, namely, the models containing group or individual perceptual overlap scores were very strongly supported over the counterpart models containing acoustic overlap scores (BFs > 150).

Table 8. Results of the mixed models for acoustic and perceptual models for groups and for individuals.

		AusE				ES			
		Est.	**SE**	**z**	***p***	**Est.**	**SE**	**z**	***p***
Perceptual Overlap Group	Intercept	1.55	0.10	14.99	<0.001	1.49	0.11	13.82	<0.001
	Effect	**−0.38**	0.06	−6.32	<0.001	**−0.49**	0.06	−7.70	<0.001
Perceptual Overlap Individual	Intercept	1.59	0.11	14.66	<0.001	1.43	0.12	12.01	<0.001
	Effect	**−0.20**	0.05	−4.25	<0.001	**−0.21**	0.05	−3.93	<0.001
Acoustic Overlap Group	Intercept	1.62	0.11	14.70	<0.001	1.49	0.12	12.26	<0.001
	Effect	**−0.36**	0.07	−5.28	<0.001	**−0.30**	0.06	−4.92	<0.001
Acoustic Overlap Individual	Intercept	1.59	0.12	13.36	<0.001	1.38	0.12	11.85	<0.001
	Effect	**−0.19**	0.06	−2.98	0.003	**−0.11**	0.04	−2.85	0.004

Next, we compared each group model, as reported in Table 8, with its counterpart individual model. For AusE listeners, the group acoustic overlap model provided modest positive support over the individual acoustic overlap model (BF = 9.97), whereas the group perceptual overlap model provided weak support over the individual perceptual overlap model (BF = 1.82). For ES listeners, both the group perceptual and acoustic overlap models provided very strong support over the individual models (BFs > 150). In summary, the pairwise model comparisons indicate that acoustic overlap scores better predict AusE listeners' discrimination performance and perceptual overlap scores better predict ES listeners' performance, and group overlap scores better predict ES listeners' performance, whereas there is less strong evidence in favour of group overlap scores predicting AusE listeners' discrimination performance.

5. General Discussion

The present study investigated whether Australian English (AusE) and European Spanish (ES) listeners differed in their categorisation and discrimination of Brazilian Portuguese (BP) vowels. Specifically, we were interested in whether acoustic similarity (based on individuals' own native production data) predicted their non-native categorisation patterns, as predicted by L2LP (Escudero 2005, 2009; van Leussen and Escudero 2015; Elvin and Escudero 2019), as well as whether perceptual similarity also predicted discrimination accuracy to a better, worse or same degree as acoustics. We further investigated whether individual native vowel production and categorisation patterns better predicted non-native discrimination than production and/or perception averages. We conducted a comprehensive acoustic analysis of the cross-linguistic differences between the listeners own native vowel production and the target BP vowels in order to predict their non-native categorisation patterns from acoustic similarities, according to L2LP principles. We further calculated the amount of acoustic and perceptual overlap (i.e., where the two vowels in a

BP contrast were acoustically similar to/perceived as the same native vowel category[ies]) in order to predict discrimination difficulty. We predicted that the greater the acoustic and/or perceptual overlap, the more difficult the BP contrast would be to discriminate.

Our results indicated that AusE and ES listeners' patterns of non-native categorisation were partially consistent with L2LP predictions based on the cross-linguistic acoustic similarity between the listeners' own native vowel productions and the target similarity. For AusE listeners, acoustic similarity successfully predicted cases of L2LP's SUBSET scenario in that each BP vowel was categorised to multiple categories of L1 vowels, as expected. For ES listeners instances of L2LP's NEW scenario, in which two L2 categories were mapped on to the same native category, were identified, also in line with acoustic predictions. Interestingly, the acoustic comparison also successfully predicted that BP /e/ and BP /o/ would be mapped to two native ES categories, which contributed to perceptual overlap.

We do find some discrepancies between our L2LP acoustic predictions and non-native categorisation patterns, particularly for the AusE listeners, in that some listener categorisation responses were not predicted by acoustic similarity. These differences are likely caused by the fact that our acoustic analysis only used F1, F2 and durational values, and did not include additional features of the vowel such as F3 (e.g., for lip rounding/vocal tract lengthening) and dynamic formant trajectories, which have been show to play an important role in AusE vowel perception and production (Elvin et al. 2016; Escudero et al. 2018; Williams et al. 2018). Another possible explanation for these differences could be the influence of orthographic labels (Escudero and Wanrooij 2010; Bassetti et al. 2015) in non-native vowel categorisation or the number of response options (Benders et al. 2012), both of which differed between the AusE and ES listeners. For instance, our acoustic analyses are not influenced by orthography, whereas listeners were presented with orthographic labels to represent each native vowel category in the non-native categorisation task. The influence of orthography on vowel perception has been demonstrated in Escudero and Wanrooij (2010) where Spanish learners of Dutch exhibited different patterns of vowel categorisation across an auditory only and auditory with orthography task (for a full review see: Escudero and Wanrooij 2010).

Turning to our results for non-native discrimination, in line with Elvin et al. (2014), the results from the present study indicate that both groups found some of the same BP contrasts easy versus difficult to discriminate. However, unlike that previous study, we found an interaction between language group and BP contrast and therefore the exact patterns and rankings of discrimination difficulty differed slightly across the two groups between these studies (see Table 9). In Elvin et al. (2014), both AusE and ES had comparable accuracy scores for BP /a/–/ɔ/, /e/–/ɛ/, /o/–/ɔ/. However, in the current study, the AusE participants found /a/–/ɔ/ to be easier than /e/–/ɛ/, which was easier than /o/–/ɔ/. For the ES listeners, BP /a/–/ɛ/ and /a/–/ɔ/ were equally the easiest to discriminate, followed by /e/–/ɛ/; and /o/–/ɔ/ was as difficult to discriminate as /i/–/e/ and /o/–/u/.

Table 9. Reported patterns of discrimination accuracy in Elvin et al. (2014) and the present study beginning from perceptually easy to perceptually difficult.

	Order of Discrimination Difficulty
Elvin et al. 2014	AusE/ES: /a/–/ɛ/>/a/–/ɔ/~/e/–/ɛ/~/o/–/ɔ/>/i/–/e/~/o/–/u/
The present study	AusE: /a/–/ɛ/>/a/–/ɔ/>/e/–/ɛ/>/o/–/ɔ/>/i/–/e/~/o/–/u/ ES: /a/–/ɛ/~/a/–/ɔ/>/e/–/ɛ/>/o/–/ɔ/~/i/–/e/~/o/–/u/

We also investigated acoustic and perceptual similarity as predictors of discrimination accuracy. The results from our generalised linear mixed models suggest that both measures are indeed reliable predictors of discrimination accuracy. Specifically, the higher the perceptual and/or acoustic similarity, the lower the accuracy scores. We ran further model comparisons in order to determine whether one measure of vowel category overlap better explained the discrimination accuracy scores for each group. Interestingly, we found that perceptual overlap scores were a better predictor of discrimination difficulty than acoustic similarity for the ES listeners. However, the opposite was true for AusE, that is, acoustic overlap scores better predicted discrimination accuracy.

The differing vowel inventory sizes could possibly explain why perceptual similarity was a better predictor of discrimination accuracy for the ES participants. Spanish has a five vowel inventory with transparent spelling, whereas Australian English has 13 vowels and opaque spelling. Therefore, in the non-native categorisation task, the ES participants are making a decision between fewer response categories than the AusE participants. Therefore, perhaps the fewer response options reduce the chance of labelling error in that task leading to strong results in the non-native categorisation task. Relatedly, data based on acoustic similarity may be better at predicting AusE discrimination accuracy because of potential labelling errors in the non-native categorisation task. In that task, AusE listeners have 13 vowels to choose from, and due to the opaque nature of English vowel spellings, the vowels were embedded in an orthographic context, which may have increased the demands and/or posed difficulties for the AusE participants. In a task with 13 response categories, there are many potential non-native categorisation patterns available, but also a greater chance of labelling errors. This may suggest that the AusE categorisation trends are not always especially strong or clear in terms of their response frequencies. This has been found in Shaw et al. (2018), where the experiment presented all English vowels with a grid of the 20 corresponding response options and AusE listeners show poor categorisation results for native Australian accented vowels. In the current study, the AusE listeners' categorisation of BP /ɔ/ is a good example of this, where the most selected option had a 32% categorisation frequency. This issue of labelling errors is not applicable in acoustic data, which may be why the acoustic analysis yields a more consistent result. To determine whether or the number of response categories and/or the opaqueness/transparency of the language influenced perception results, future studies could compare languages that have the same number of vowels, but differ in terms of the degree of orthographic transparency.

We also investigated whether measures of acoustic and perceptual categorisation overlap are equally good predictors of discrimination accuracy at both group and individual levels (i.e., using individual overlap scores vs. group averages). This was made possible by our inclusion of native production data. The results from our analyses revealed that both measures do in fact predict discrimination accuracy. However, we found that the model based on group averages was a better predictor of discrimination accuracy than individual score averages for the ES group. Yet the evidence for group scores was not as strong for predicting the AusE listeners' discrimination performance. This finding goes against the L2LP model claims as well as studies that show the importance of individual variation in predicting L2 perceptual development. It is possible that the group model provides a better estimate of individual behaviour because the averages across the individuals are less noisy than the individual averages. That is, population data are less affected by response errors/variability in responses when doing a non-native categorisation task, as it aggregates responses from many trials (compared to a single listener with fewer trials), weakening the influence of any "outlier" behaviour. Finally, the fact that there is not actually much difference between individual and group data for the AusE group suggests that assimilation/similarity patterns at individual and group level are not as reliable or are quite variable (because there are so many assimilation possibilities available). Further investigation into the effect of individual differences for L2 development is required.

In sum, our findings indicate that listeners' non-native categorisation patterns are largely predicted by a detailed acoustic comparison of the native and target languages, with data collected from the same populations for both vowel productions and perceptual testing, mostly in keeping with L2LP predictions. Importantly, we find that AusE listeners do not have an advantage when perceiving non-native vowels despite their native language

(English) having a larger and more complex vowel inventory than that of the ES listeners (Spanish). In fact, we find that listeners' discrimination patterns are largely dependent on the L2LP learning scenario identified for each vowel contrast, which were similar across the two language groups. That is, contrasts which contained evidence of L2LP's NEW or SUBSET scenarios (containing an acoustic or perceptual overlap where the two vowels in a given BP contrast are acoustically similar to, or categorised to, the same native category[s]) resulted in similar discrimination difficulties for ES and AusE listeners, as both scenarios resulted in a failure to detect a distinction between the non-native vowels. In addition, both of these learning scenarios are likely to be more difficult than contrasts where only the SIMILAR scenario is present. These findings are also consistent with previous studies (Bohn et al. 2011; Levy 2009; Tyler et al. 2014; Best et al. 2019) testing PAM's theoretical predictions that a higher degree of perceived phonetic similarity (i.e., perceptual overlap) between members of a non-native contrast, as observed by perceptual assimilation patterns, is associated with a greater level of discrimination difficulty for that contrast.

We further found that performance in non-native discrimination can be predicted by measures of acoustic and perceptual similarity using both group and individual data, although we found that perceptual similarity was a better predictor for ES and acoustic similarity for AusE. We also found that group data better explained ES discrimination accuracy, but not as clearly so for AusE listeners. We suspect that these findings are related to vowel inventory size differences between their native languages, and the nature of the response categories in the non-native categorisation task. At this stage, it is difficult to do a direct comparison between the overlap measures (acoustic vs. perceptual and group vs. individual) because both predict the discrimination data. Further studies on acoustic predictions should include F3 and dynamic vowel measures in addition to F1 and F2 static and durational measures, in line with previous studies showing the importance of dynamic information for AusE vowels in particular (Elvin et al. 2016; Escudero et al. 2018; Williams et al. 2018). It seems that the nature of the non-native categorisation task and the differing language backgrounds complicate any conclusions that could be drawn in regards to which measure is better. For now, it is sufficient to state that both acoustic and perceptual similarity predict performance on discrimination and further investigation would be required to identify whether one measure is better than the other, or perhaps it is a case where different measures are more suited to different languages.

Finally, according to the L2LP model, perception is linked to spoken word recognition and production, thus future studies should compare data from the tasks of the present study with the same listeners' spoken word recognition and non-native vowel production data. Indeed, some prior studies (e.g., Broersma 2002; Escudero et al. 2013; Escudero et al. 2008; Pallier et al. 2001; Weber and Cutler 2004) have shown that vowel contrasts that are difficult to perceive are also difficult in spoken word learning and word recognition tasks. Therefore, we would expect that these same listeners' patterns of discrimination difficulty could be used to predict difficulty in a spoken word learning and word recognition task containing the same pseudo-words used in the present study. Furthermore, the findings from the present study could also apply to non-native speech production. In particular, the SLM and L2LP theoretical models claim that listeners' production of non-native or L2 sounds is influenced by their perception of these sounds in the L1. However, to date, most studies (e.g., Diaz Granado 2011; Flege et al. 1999) have used the theoretical framework of SLM to test L2 production. Considering that the L2LP model posits a direct link between non-native and L2 production, it is perhaps surprising that very few studies have tested this claim (e.g., Rauber et al. 2005). Thus, acoustic analyses and in particular the acoustic and perceptual overlap scores could be used to predict the patterns in the same listeners' non-native vowel productions. In conclusion, future research is required that adequately tests the L2LP model predictions for the role of perception in both word recognition and non-native production, and compares them to other models of non-native and L2 speech perception and production.

Author Contributions: J.E., P.E., J.A.S. and C.T.B. conceived and designed the experiments. J.E. ran all experiments. J.E. and D.W. conducted all the analyses. J.E. wrote the original draft. J.E., D.W., J.A.S., C.T.B., and P.E. reviewed and edited this manuscript. All authors have read and agreed to the published version of this manuscript.

Funding: This work was funded by an Australian Postgraduate Award, the MARCS Research Training Scheme and the Australian Research Council (ARC) Centre of Excellence for the Dynamics of Language (CE140100041). PE's work was supported by an ARC Future Fellowship (FT160100514).

Institutional Review Board Statement: The study was conducted according to the guidelines of the National Ethics Application Form, and approved by the Human Research Ethics Committee) WESTERN SYDNEY UNIVERSITY (H10427 approved 01/11/2013).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

Alispahic, S., K. E. Mulak, and P. Escudero. 2017. Acoustic properties predict perception of unfamiliar Dutch vowels by adult Australian English and Peruvian Spanish listeners. *Frontiers in Psychology* 8: 52. [CrossRef]

Alispahic, S., P. Escudero, and K. E. Mulak. 2014. Difficulty in discriminating non-native vowels: Are Dutch vowels easier for Australian English than Spanish listeners? Paper presented at Annual Conference of the International Speech Communication Association, INTERSPEECH, (SEPTEMBER), Singapore, September 14–18; pp. 1293–96. [CrossRef]

Antoniou, M., and P. C. Wong. 2015. Poor phonetic perceivers are affected by cognitive load when resolving talker variability. *The Journal of the Acoustical Society of America* 138: 571–74. [CrossRef]

Bassetti, B., P. Escudero, and R. Hayes-Harb. 2015. Second language phonology at the interface between acoustic and orthographic input. *Applied Psycholinguistics* 36: 1–6. [CrossRef]

Benders, T., P. Escudero, and M. J. Sjerps. 2012. The interrelation between acoustic context effects and available response categories in speech sound categorisation. *The Journal of the Acoustical Society of America* 131: 3079–87. [CrossRef]

Best, C. T. 1994. The Emergence of Native-Language Phonological Influences in Infants: A Perceptual Assimilation Model. *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words* 167: 167–224.

Best, C. T. 1995. A direct realist perspective on cross-language speech perception. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Edited by W. Strange. Timonium: York Press, pp. 171–204.

Best, C. T., and M. D. Tyler. 2007. Non-native and second-language speech perception: Commonalities and complementarities. In *Language Experience in Second-Language Speech Learning: In Honor of James Emil Flege*. Edited by O. Bohn and M. J. Munro. Amsterdam: John Benjamins, pp. 13–34.

Best, C. T., C. Avesani, M. D. Tyler, and M. Vayra. 2019. Perceiving non-native consonants that utilize articulators not employed contrastively in the native language: Revisiting PAM and the Articulatory Organ Hypothesis. In *A Sound Approach to Language Matters: In Honor of Ocke-Schwen Bohn*. Edited by A. M. Nyvad, M. Hejná , A. Højen, A. B. Jespersen and M. H. Sørensen. Aarhus: Aarhus University Press, pp. 13–40. [CrossRef]

Blackwood Ximenes, A., J. A. Shaw, and C. Carignan. 2017. A comparison of acoustic and articulatory methods for analyzing vowel differences across dialects: Data from American and Australian English. *The Journal of the Acoustical Society of America* 142: 363–77. [CrossRef]

Bohn, O. S., C. T. Best, C. Avesani, and M. Vayra. 2011. Perceiving through the lens of native phonetics: Italian and Danish listenerś perception of English consonant contrasts. Paper presented at XVII International Congress of Phonetic Sciences, Hong Kong, China, August 17–21; pp. 336–39.

Bohn, O., and J. E. Flege. 1990. Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics* 11: 303–28. [CrossRef]

Broersma, M. 2002. Comprehension of non-native speech: Inaccurate phoneme processing and activation of lexical competitors. Paper presented at the Seventh International Conference on Spoken Language Processing, Denver, CO, USA, September 16–20; pp. 261–64.

Brysbaert, M., and M. Stevens. 2018. Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition* 1: 9. [CrossRef]

Chládková, Kateřina, and Paola Escudero. 2012. Comparing vowel perception and production in Spanish and Portuguese: European versus Latin American dialects. *The Journal of the Acoustical Society of America* 131: EL119–EL125. [CrossRef] [PubMed]

Colantoni, L., J. Steele, and P. Escudero. 2015. *Second Language Speech*. Cambridge: Cambridge Univ Press.

Cox, F. 2006. The Acoustic Characteristics of/hVd/Vowels in the Speech of some Australian Teenagers. *Australian Journal of Linguistics* 26: 147–79. [CrossRef]

Curtin, S., C. Fennell, and P. Escudero. 2009. Weighting of vowel cues explains patterns of word—Object associative learning. *Developmental Science* 12: 725–31. [CrossRef]

Diaz Granado, M. 2011. L2 and L3 Acquisition of the Portuguese Stressed Vowel Inventory by Native Speakers of English. Ph.D. disstertation, The University of Arizona, Tucson, AZ, USA.

Díaz, B., H. Mitterer, M. Broersma, and N. Sebastian-Gallés. 2012. Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access. *Learning and Individual Differences* 22: 680–89. [CrossRef]

Elvin, J. 2016. The Role of the Native Language in Non-Native Perception and Spoken Word Recognition: English vs. Spanish Learners of Portuguese. Ph.D. dissertation, Western Sydney University, Sydney, NSW, Australia.

Elvin, J., and P. Escudero. 2019. Cross-Linguistic Influence in Second Language Speech: Implications for Learning and Teaching. In *Cross-Linguistic Influence: From Empirical Evidence to Classroom Practice*. Cham: Springer, pp. 1–20.

Elvin, J., D. Williams, and P. Escudero. 2016. Dynamic acoustic properties of monophthongs and diphthongs in Western Sydney Australian English. *The Journal of the Acoustical Society of America* 140: 576–81. [CrossRef] [PubMed]

Elvin, J., P. Escudero, and P. Vasiliev. 2014. Spanish is better than English for discriminating Portuguese vowels: Acoustic similarity versus vowel inventory. *Frontiers in Psychology* 5: 1–8. [CrossRef]

Escudero, P. 2001. The role of the input in the development of L1 and L2 sound contrasts: Language-specific cue weighting for vowels. In *Proceedings of the 25th Annual Boston University Conference on Language Development*. Somerville: Cascadilla Press, pp. 250–61.

Escudero, P. 2005. *Linguistic Perception and Second Language Acquisition: Explaining the Attainment of Optimal Phonological Categorization*. Amsterdam: Netherlands Graduate School of Linguistics.

Escudero, P. 2009. The linguistic perception of similar L2 sounds. *Phonology in Perception* 15: 152–90.

Escudero, P., and D. Williams. 2011. Perceptual assimilation of Dutch vowels by Peruvian Spanish listeners. *The Journal of the Acoustical Society of America* 129: EL1–7. [CrossRef]

Escudero, P., and D. Williams. 2012. Native dialect influences second-language vowel perception: Peruvian versus Iberian Spanish learners of Dutch. *The Journal of the Acoustical Society of America* 131: EL406–12. [CrossRef]

Escudero, P., and K. Chládková. 2010. Spanish listeners' perception of American and Southern British English vowels. *The Journal of the Acoustical Society of America* 128: EL254–EL259. [CrossRef]

Escudero, P., and K. Wanrooij. 2010. The Effect of L1 Orthography on Non-native Vowel Perception. *Language and Speech* 53: 343–65. [CrossRef]

Escudero, P., and P. Boersma. 2002. The Subset Problem in L2 Perceptual Development: Multiple-Category Assimilation by Dutch Learners of Spanish. Paper presented at 26th annual Boston University conference on language development, Boston, MA, USA, November 5–7; pp. 208–19.

Escudero, P., and P. Boersma. 2004. Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26: 551–85. [CrossRef]

Escudero, P., and P. Vasiliev. 2011. Cross-language acoustic similarity predicts perceptual assimilation of Canadian English and Canadian French vowels. *The Journal of the Acoustical Society of America* 130: EL277–83. [CrossRef]

Escudero, P., and R. A. H. Bion. 2007. Modeling Vowel Normalization and Sound Perception as Sequential Processes. Available online: http://icphs2007.de/conference/Papers/1313/1313.pdf (accessed on 22 February 2021).

Escudero, P., B. Sisinni, and M. Grimaldi. 2014. The effect of vowel inventory and acoustic properties in Salento Italian learners of Southern British English vowels. *The Journal of the Acoustical Society of America* 135: 1577–84. [CrossRef] [PubMed]

Escudero, P., K. E. Mulak, J. Elvin, and N. M. Traynor. 2018. "Mummy, keep it steady": Phonetic variation shapes word learning at 15 and 17 months. *Developmental Science* 21: e12640. [CrossRef]

Escudero, P., M. Broersma, and E. Simon. 2013. Learning words in a third language: Effects of vowel inventory and language proficiency. *Language and Cognitive Processes*, 1–16. [CrossRef]

Escudero, P., P. Boersma, A. S. Rauber, and H. R. Bion. 2009. A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *The Journal of the Acoustical Society of America* 126: 1379–93. [CrossRef] [PubMed]

Escudero, P., R. Hayes-Harb, and H. Mitterer. 2008. Novel second-language words and asymmetric lexical access. *Journal of Phonetics* 36: 345–60. [CrossRef]

Faris, M. M., C. T. Best, and M. D. Tyler. 2016. An examination of the different ways that non-native phones may be perceptually assimilated as uncategorized. *The Journal of the Acoustical Society of America* 139: 1–5. [CrossRef] [PubMed]

Flege, J. E. 1995. Second language speech learning: Theory, findings, and problems. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Timonium: York Press, pp. 233–76.

Flege, J. E., I. R. A. Mackay, and D. Meador. 1999. Native Italian speakers ' perception and production of English vowels. *The Journal of the Acoustical Society of America* 106: 2973–87. [CrossRef] [PubMed]

Flege, J. E., M. J. Munro, and I. R. A. Mackay. 1995. Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America* 97: 3125. [CrossRef] [PubMed]

Flege, J. E., O.-S. Bohn, and S. Jang. 1997. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics* 25: 437–70. [CrossRef]

Gilichinskaya, Y. D., and W. Strange. 2010. Perceptual assimilation of American English vowels by inexperienced Russian listeners. *The Journal of the Acoustical Society of America* 128: EL80–5. [CrossRef] [PubMed]

Gordon, L. S. 2008. *Factors Affecting English Speakers' Perception of L2 Vowels. ProQuest*. Washington: Georgetown University.

Green, P., and C. J. MacLeod. 2016. SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution* 7: 493–98. [CrossRef]

Iverson, P., and B. G. Evans. 2007. Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration. *The Journal of the Acoustical Society of America* 122: 2842–54. [CrossRef] [PubMed]

Iverson, P., and B. G. Evans. 2009. Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America* 126: 866–77. [CrossRef]

Jin, Y., B. Díaz, M. Colomer, and N. Sebastián-Gallés. 2014. Oscillation encoding of individual differences in speech perception. *PLoS ONE* 9. [CrossRef] [PubMed]

Kisler, T., F. Schiel, and H. Sloetjes. 2012. Signal processing via webservices: The use case WebMAUS. Paper presented at Digital Humanities, Hamburg, Germany, July 16–22; pp. 30–34.

Levy, E. S. 2009. On the assimilation-discrimination relationship in American English adults' French vowel learning. *The Journal of the Acoustical Society of America* 126: 2670–82. [CrossRef]

Levy, E. S., and F. F. Law. 2010. Production of French vowels by American-English learners of French: Language experience, consonantal context, and the perception-production relationship. *The Journal of the Acoustical Society of America* 128: 1290–305. [CrossRef] [PubMed]

Lobanov, B. M. 1971. Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America* 49: 606–8. [CrossRef]

Mayr, R., and P. Escudero. 2010. Explaining individual variation in L2 perception: Rounded vowels in English learners of German. *Bilingualism: Language and Cognition* 13: 279–97. [CrossRef]

Morrison, G. S. 2003. Perception and Production of Spanish Vowels by English Speakers. Paper presented at 15th International Congress of Phonetic Sciences, Barcelona, Spain, August 3–9; vol. 2003, pp. 1533–36.

Morrison, G. S. 2009. L1-Spanish Speakers' Acquisition of the English/i/–//Contrast II: Perception of Vowel Inherent Spectral Change1. *Language and Speech* 52: 437–62. [CrossRef]

Noiray, A., K. Iskarous, and D. H. Whalen. 2014. Variability in English vowels is comparable in articulation and acoustics. *Laboratory phonology* 5: 271. [CrossRef] [PubMed]

Pallier, C., A. Colomé, and N. Sebastiáń Gallés. 2001. The influence of native-language phonology on lexical access: Exemplar-based vs. abstract lexical entries. *Pyschological Science* 12: 445–49. [CrossRef] [PubMed]

Rauber, A. S., P. Escudero, R. A. H. Bion, and B. O. Baptista. 2005. The Interrelation between the Perception and Production of English Vowels by Native Speakers of Brazilian Portuguese Graduate Program in Applied Linguistics. *Interspeech* 2: 2913–16.

Sebastián-Gallés, N., and B. Díaz. 2012. First and Second Language Speech Perception: Graded Learning. *Language Learning* 62: 131–47. [CrossRef]

Shaw, J. A., C. T. Best, G. Docherty, B. Evans, P. Foulkes, and J. Hay. 2018. Resilience of English vowel perception across regional accent variation. *Laboratory Phonology* 9: 11. [CrossRef]

Smith, B. L., and R. Hayes-Harb. 2011. Individual differences in the perception of final consonant voicing among native and non-native speakers of English. *Journal of Phonetics* 39: 115–20. [CrossRef]

Strange, W., O.-S. Bohn, S. Trent, and K. Nishi. 2004. Acoustic and perceptual similarity of North German and American English vowels. *The Journal of the Acoustical Society of America* 115: 1791. [CrossRef]

Thomas, E. R., and T. Kendall. 2007. NORM: The Vowel Normalization and Plotting Suite: An Online Tool for Sociophonetic Vowel Normalization. Available online: http://lingtools.uoregon.edu/norm/ (accessed on 22 February 2021).

Tyler, M. D., C. T. Best, A. Faber, and A. G. Levitt. 2014. Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica* 71: 4–21. [CrossRef] [PubMed]

van Leussen, J.-W., and P. Escudero. 2015. Learning to perceive and recognize a second language: The L2LP model revised. *Frontiers in Psychology* 6: 1–12. [CrossRef]

Vasiliev, P. 2013. *The Initial State for Californian English Learners of Spanish and Portuguese Vowels*. Los Angeles: University of California.

Wagenmakers, E.-J. 2007. A practical solution to the perverse problems of p-values. *Psychonomic Bulletin and Review* 14: 779–804. [CrossRef]

Wanrooij, K., P. Escudero, and M. E. Raijmakers. 2013. What do listeners learn from exposure to a vowel distribution? An analysis of listening strategies in distributional learning. *Journal of Phonetics* 41: 307–19. [CrossRef]

Watson, C. I., and J. Harrington. 1999. Acoustic evidence for dynamic formant trajectories in Australian English vowels. *The Journal of the Acoustical Society of America* 106: 458–68. [CrossRef] [PubMed]

Weber, A., and A. Cutler. 2004. Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language* 50: 1–25. [CrossRef]

Whalen, D. H., W. R. Chen, M. K. Tiede, and H. Nam. 2018. Variability of articulator positions and formants across nine english vowels. *Journal of phonetics* 68: 1–14. [CrossRef]

Williams, D., and P. Escudero. 2014. A cross-dialectal acoustic comparison of vowels in Northern and Southern British English. *The Journal of the Acoustical Society of America* 136: 2751–61. [CrossRef] [PubMed]

Williams, D., P. Escudero, and A. Gafos. 2018. Spectral change and duration as cues in Australian English listeners' front vowel categorisation. *The Journal of the Acoustical Society of America* 144: EL215–21. [CrossRef]

Yazawa, K., J. Whang, M. Kondo, and P. Escudero. 2020. Language-dependent cue weighting: An investigation of perception modes in L2 learning. *Second Language Research* 36: 557–81. [CrossRef]

Yu, A. C., and G. Zellou. 2019. Individual differences in language processing: Phonology. *Annual Review of Linguistics* 5: 131–50. [CrossRef]

languages

MDPI

Article

Cross-Linguistic Interactions in Third Language Acquisition: Evidence from Multi-Feature Analysis of Speech Perception

Magdalena Wrembel [1,*], Ulrike Gut [2], Romana Kopečková [2] and Anna Balas [1]

[1] Faculty of English, Adam Mickiewicz University, Wieniawskiego 1, 61-712 Poznań, Poland; abalas@wa.amu.edu.pl

[2] English Department, University of Münster, Schlossplatz 2, 48149 Münster, Germany; gut@uni-muenster.de (U.G.); kopeckor@tcd.ie (R.K.)

* Correspondence: magdala@wa.amu.edu.pl

Received: 1 September 2020; Accepted: 28 October 2020; Published: 3 November 2020

Abstract: Research on third language (L3) phonological acquisition has shown that Cross-Linguistic Influence (CLI) plays a role not only in forming the newly acquired language but also in reshaping the previously established ones. Only a few studies to date have examined cross-linguistic effects in the speech perception of multilingual learners. The aim of this study is to explore the development of speech perception in young multilinguals' non-native languages (L2 and L3) and to trace the patterns of CLI between their phonological subsystems over time. The participants were 13 L1 Polish speakers (aged 12–13), learning English as L2 and German as L3. They performed a forced-choice goodness task in L2 and L3 to test their perception of rhotics and final obstruent (de)voicing. Response accuracy and reaction times were recorded for analyses at two testing times. The results indicate that CLI in perceptual development is feature-dependent with relative stability evidenced for L2 rhotics, reverse trends for L3 rhotics, and no significant development for L2/L3 (de)voicing. We also found that the source of CLI differed across the speakers' languages: the perception accuracy of rhotics differed significantly with respect to stimulus properties, that is, whether they were L1-, L2-, or L3-accented.

Keywords: multilingualism; third language acquisition; speech perception; rhotics; final obstruent devoicing

---

1. Introduction-Bilingual vs. Multilingual Perspective

In this contribution, we explore cross-linguistic interactions between phonological subsystems in third language acquisition, based on evidence from multifeature analysis of speech perception. Research on third language (L3) phonological acquisition has shown that Cross-Linguistic Influence (CLI) plays a role not only in forming the newly acquired language but also in reshaping the previously established ones (cf. Wrembel and Cabrelli 2018). There is scarcity of evidence from perceptual studies, however, which seems unfortunate considering that speech perception has been seen as driving the process of non-native phonological acquisition. The most influential second language (L2) phonology models use (cross-language phonetic) perception to explain the outcomes of L2 speech learning e.g., Speech Learning Model, (Flege 1995; Flege and Bohn 2020) and Perceptual Assimilation Model (Best 1995; Tyler 2019). It would, therefore, seem beneficial and necessary that L3 phonology research complements its findings by examining cross-linguistic interactions in multilingual perceivers in order to be ultimately in a more favorable position to both explicate their production and gain a more complete picture of multilingual phonological acquisition.

Third language acquisition (TLA) has recently gained recognition as an independent field of enquiry from second language acquisition (SLA). Scholars working on this new perspective maintain that the former is inherently more complex than the latter, as it involves a quality change in the language learning and processing (e.g., Cenoz et al. 2001; De Angelis 2007). They imply that the process of learning the first foreign language (L2) is fundamentally different from the process of learning a third or additional language (L3/Ln), mainly because of enhanced language awareness, language learning strategies, and increased potential for cross-linguistic interactions between L1, L2, L3, or Ln that occur in additional language acquisition. A number of linguistic and psycholinguistic studies support these claims by providing evidence for the existence of qualitative and quantitative differences in processing the third language as compared to the first or second language (Cenoz and Jessner 2000; Cenoz et al. 2001; Hufeisen and Lindemann 1997). From a theoretical linguistic perspective, Flynn et al. (2004) argue that the study of L3 acquisition can offer new insights into the process of language learning that exceed those offered by investigations of the first or the second language.

One of the major differences between the acquisition of a second and a third language is that L3 learners have already acquired their first foreign language, and, thus, they can resort to some conscious linguistic knowledge as well as language-learning experience and strategies (cf. De Angelis 2007). Multilingual learners, thus, have at their disposal a broadened phonetic repertoire, a raised level of metalinguistic awareness, and potentially enhanced perceptual sensitivity, which may facilitate the learning of a subsequent phonological system (cf. Gut 2010; Wrembel 2015). In a dedicated volume on "Universal or diverse paths to English phonology", Gut et al. (2015) attempt a comprehensive comparison between the acquisition of phonology from a SLA vs. TLA perspective, showing that L3 learners' development of perception and production differs sharply from that of L2 learners' in being more differentiated and constrained by a greater number of factors.

The extant findings from L3 phonology research suggest that any of the previously or currently acquired languages can serve as a source for CLI in the perception and production of target segments and suprasegmentals, and that this phenomenon is multidirectional (cf. Cabrelli and Wrembel 2016). We have a growing understanding of the combination of factors conditioning the different types of phonological CLI in L3 learning, such as proficiency in the respective languages, (psycho)typology as well as the type of phonological task performed (for an overview, see Wunder 2014). However, so far the investigations have been mostly limited to a single feature and/or one testing time, thus exploring this question with more phonetic features and longitudinally seems paramount for our understanding of the relative effect of cross-linguistic processes in non-native speech learning, and speech perception in particular.

In the present paper, we examine L2 and L3 speech perception of two phonetic features, which have a different standing in the phonological repertoire of the multilinguals of this study, over the course of the first year of their instructed L3 learning. We seek to investigate how and to what extent phonological CLI may change over time in multilingual perceivers.

2. Non-Native Speech Perception

Considering that only a few studies to date have examined speech perception of multilingual learners (cf. Balas et al. 2019; Wrembel et al. 2019; Nelson 2020), models of L2 speech perception may serve as an informative starting point for the formulation of predictions for L3 learners, taking into consideration the learners' enlarged phonological repertoire as well as greater language learning experience. Most L2 speech perception models have predicted accuracy of perception on the basis of similarities and differences between L1 and L2 sounds. Starting with Lado (1957) Contrastive Analysis Hypothesis, L2 phonemes that are similar to L1 phonemes were considered easy to perceive and L2 sounds that are different from the L1 sounds difficult. Eckman (1977) Markedness Differential Hypothesis proposed that target language structures that are both different and more marked should prove difficult for learners, whereas structures that are different but less marked should not pose difficulties. The Speech Learning Model (SLM; Flege 1995) predicts that it is the fairly similar L2 sounds (to their L1 counterparts) that are most challenging for L2 learners to acquire, as they are subject to equivalence classification, i.e., they are perceptually equated with existing L1 categories.

Conversely, the sounds that do not resemble any of the L1 categories may enhance the process of category formation, and, hence, be perceived accurately. Similarly, the Perceptual Assimilation Model (PAM; Best 1995; Best and Tyler 2007) presupposes that not all target language sounds are equally challenging for learners, but it focuses on non-native contrasts rather than on individual phonemes. Discrimination of non-native sounds varies depending on how a non-native contrast is assimilated and goodness-rated to native language phonological categories, resulting in at least four different assimilation patterns for each non-native sound contrast (Best 1995, pp. 194–98).

Most relevantly for the present study, PAM predicts a continuous refinement of L2 learners' speech perception as a function of their extended experience with learning the L2 (PAM-L2; Best and Tyler 2007). With time, learners are likely to enjoy not only more L2 input but also to gain greater experience in producing the target contrasts and to increase their knowledge of L2 (minimal pair) vocabulary (Bundgaard-Nielsen et al. 2011). According to the model, L2 learners are, thus, expected to start perceiving within-category differences and develop new categories for the non-native sounds and contrasts. The way this category refinement may reshape in the context of L3 learning, particularly when L2 continues to develop too is still to be examined (for the first attempt, see Wrembel et al. 2019).

As non-native speech perception is characterized by considerable inter-listener variation, the Second Language Linguistic Perception Model (L2LPM; Escudero and Boersma 2004; Escudero 2005, 2009) concentrates on individual developmental paths on the basis of a detailed acoustic comparison of the production of L1 and L2 sounds. Two main learning scenarios are present for L2 learners, according to this model: When two L2 sounds are categorized to the same native language category, the learner needs to create a new category for one of the L2 sounds or split the existing category. When two L2 sounds are heard as separate L1 categories, the learner's task is to shift category boundaries to accommodate the L2 sounds. The latter scenario in which an L2 sound is perceived as more than one native category may be challenging as it may lead to overdifferentiation in the L2. The speed of perceptual learning in this model is, thus, predicted to depend on the particular learning scenario and richness of both L1 and L2 input that an individual learner enjoys in their learning environment.

2.1. Development of Non-Native Speech Perception

Previous research on the role of experience in the perception of non-native sounds and contrasts has yielded mixed results. Flege (1991); Baker et al. (2002); Kopečková (2012), and Rallo Fabra and Romero (2012) reported (immersion) experience effects on the discrimination and identification of at least some L2 English vowels and consonants of speakers of diverse L1 backgrounds, Cebrian (2006) found no significant differences between experienced and inexperienced Catalan-Spanish bilinguals in categorizing English /iː/ and /ɪ/ vowels. The former group of English learners had resided in Canada for an average of 25 years, while the latter group consisted of undergraduate students of English philology living in Barcelona. Cebrian (2006) reported both learner groups to rely on duration rather than spectral cues in the perception of the target contrast. In a similar vein, Broesma (2005) showed that highly experienced Dutch learners of English can accurately categorize word-final lenis-fortis contrasts, but do not use native-like weighing of cues for voicedness for this familiar contrast (present in Dutch) in an unfamiliar coda position.

Mixed findings have also been reported in perception training studies. For instance, Bradlow et al. (1999) found a long-term increase in identification accuracy of English liquids by L1 Japanese speakers. Anderson (2011) showed in a study with American English learners of Spanish that after about three weeks of identification training, some of the learners perceived the Spanish tap-trill contrast highly variably first, but then it perceptually stabilized with time; that some perceived the acoustic differences rather well in the beginning, but also revealed little change and no bifurcation of the existing phoneme category, and finally that there were also "non-learners" who showed no progress in the perception of this novel contrast. The question of refinement of non-native categories for diverse phoneme types and most crucially, under what type of learning experience it happens, thus remains at present unanswered.

2.2. Previous L3 Speech Perception Studies

As argued in previous sections, one type of learning experience that may offer important insights into the process of phonological learning in general and cross-linguistic interaction in particular is that of additional/L3 phonological learning. In one of the first studies examining phonological CLI in L3 acquisition, Wrembel et al. (2019) showed that beginner L3 Polish learners perceptually assimilate L3 sibilants to both their L1 German and L2 English categories, with preference for the latter. They can perceive subtle differences between highly similar vowel sounds across the three languages and seem to develop separate L3 categories for them. Beginner L3 learners were, thus, theorized in this study to behave similarly to experienced L2 learners thanks to their extended prior linguistic and learning experience. These are important initial insights, yet longitudinal studies examining the development of speech perception beyond only the L3 are needed to gain a more holistic picture of cross-linguistic mapping processes in multilingual learners, and possible changes thereof over time.

Some first attempts for this methodologically challenging endeavor appeared in Balas et al. (2019) and Nelson (2020). Although an examination of category formation in multilingual speech perception was the main aim of neither of these longitudinal studies, the reported findings into the development of L2 and L3 perception jointly shed at least some light on the process. In a study that stems from the same research project as the present paper, Balas et al. (2019) examined the perception of L2 and L3 rhotic sounds in two groups of young multilinguals five and nine months into their first year of L3 learning. Both L1 Polish and L1 German speakers were found to perceive L2 English rhotics highly accurately and consistently after about five years of learning the language, suggesting fairly stable phonetic categories for this novel sound (in relation to their L1) and no perceptual change as a result of the one year of additional language learning. L1 German speakers were further found to perceive the novel L3 Polish alveolar trills and taps highly accurately, and significantly better and more consistently than L1 Polish speakers did in perceiving L3 German uvular fricatives; the accuracy in perceiving the novel sound further dropped significantly between the two testing times for the latter learner group. The findings were interpreted as suggesting a joint effect of the learner's L1, but not L2, markedness and L2/L3 proficiency in the perception of rhotic sounds by multilingual learners. The present contribution expands on and refocuses this study.

Nelson (2020) examined young and adult L3 learners' perception of the /v-w/ contrast, present in their L2 but not L1, reporting more accurate and faster discrimination ability in the L3 than in the L2 after only a few hours of L3 input. The author hypothesized a positive 'novelty effect' for the L3 learners, maintaining that very initial learners may not automatically assimilate novel sounds to their pre-existing categories (whether those of L1 or L2) but rather resource acoustic cues available to them and tap possible yet different processing and phonological skills at that stage of L3 phonological learning. With respect to their L2 perception development, the young learners evidenced a drop in accuracy after around 10 weeks of their L3 learning, which was interpreted as suggesting a reverse cross-linguistic effect in the form of a temporary 'perceptual confusion'. However, after ten months of learning the L3, the novelty effect as well as the negative cross-linguistic effect disappeared for the young L3 learners, who perceived the contrast in their L2 and L3 similarly (67% and 74% accuracy levels).

To sum up, a common denominator for the existing L3 perception studies is that the phonological space of multilinguals seems to be reshaped relatively early in the course of learning the new L3, and that category boundaries can be expanded to accommodate L1, L2, and L3 categories of similar phonetic types, while new L3 categories for novel phonetic types may be formed. Initial sensitivity to phonetic contrasts may also deteriorate with time as a result of language interactions and be modulated by the status of various contrasts in L3 acquisition, including that of markedness. In the present paper, we attempt to contribute to these emerging findings by examining the perception of novel rhotic sounds (both in the L2 and L3 of the multilinguals, and more marked in their L3) and the perception of final obstruent (de)voicing (more marked in their L2) in the first months of L3 learning.

3. The Present Study

The aim of this study is to explore the development of speech perception in young multilinguals' non-native languages (L2 and L3) and to trace the patterns of cross-linguistic mappings over the first year of L3 learning. This study forms a part of the international MULTI-PHON research project, in which speech perception and production was investigated with a battery of tests in two parallel groups of young adolescents in Polish and German schools.

3.1. Participants

The participants were 13 L1 Polish speakers (aged 12–13) who had been learning English as their L2 at school for five years (pre-intermediate level) and who had just started to learn German as their L3 in an instructed setting. They were observed over the first year of L3 learning. Our strict inclusion criteria featured no prior command of German, only Polish as an L1, no additional languages, and data availability at all testing times, thus, for the sake of the present analysis the number of participants was reduced from a larger participant pool (initially 24) to 13 speakers with a homogeneous profile (see Table 1).

Table 1. Participant profiles.

	Mean	*SD*
Age (years)	12.25	0.41
AOL2 (age of onset of L2 English)	5.50	0.80
AOL3 (age of onset of L3 German)	12.25	0.62
Hrs of L2 instruction per week	3	--
Hrs of L3 instruction per week	5	--
Self-evaluation in L2 *	3.65	0.51
Self-evaluation in L3 *	3.33	0.58
Female/male ratio	8/5	--

* Self-evaluation of proficiency was assessed on a 5-point scale (1 = very poor, 5 = very good).

An informed consent was obtained from all the subjects who participated in the study, their parents, and the school authorities where the data was collected. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ministry of Education in Brandenburg on 17/07/2017 (ref. number 51/2017).

Language background interviews were conducted in the participants' L1 Polish at the very onset of the project in order to collect information about the individual learner's language backgrounds, including information about their language learning history (i.e., age of learning, length and intensity of instruction), language use (declared percentage in varied situations/contexts), self-evaluation of proficiency (at the onset of instructed L3 learning), and attitudes towards foreign language learning.

3.2. Features under Investigation

Two phonetic features were selected for investigation, rhotics and final obstruent devoicing, since they have a relatively different standing in the phonological repertoire of the L3 learners in this study (see Table 2). The former sounds are realized differently in each of the speakers' three languages, whereas the latter process is productive in their L1 and L3 but not L2.

Table 2. Selected features under analysis in the present study.

Language	**Rhotics**	**Obstruent Devoicing in Syllable-Final Position**
Polish	/r/	Yes
English	/ɹ/	No
German	/ʀ/ and /ʁ/	Yes

3.2.1. Rhotics

In spite of belonging to a phonological natural class, for which there are more phonological than phonetic arguments (cf. Ladefoged and Maddieson 1996), rhotics exhibit large interlanguage variability. In the three languages under investigation in this paper, the distribution of rhotics is as follows: Polish has the alveolar trill, which may be produced as a tap intervocalically or in fast speech (Jassem 2003). In standard German, the conservative uvular trill /ʀ/, occurring in word-initial or in stressed positions, is usually produced as the uvular fricative /ʁ/ (Kohler 1999). English rhotics include British English postalveolar approximant /ɹ/ and prevocalically [/ɹ/], and an American English retroflex approximant (Ladefoged and Maddieson 1996) articulated either with tongue retroflexion or bunching (Ladefoged 2001). The English rhotic is generally voiced except when adjacent to a voiceless obstruent. It occurs in syllable-initial (e.g., run/ɹʌn/), and syllable-final position (e.g., poor/pɔɹ/) (not in British English), both as singletons and in clusters (e.g., tree/tɹi/; heart /haɹt/). Both English rhotic sounds are continuants, as opposed to the 'interrupted' variants such as taps or trills in Polish. Worthy of note is the fact that in all three languages, the rhotic sounds are represented orthographically using the <r> letter. This suggests that orthography may promote multiple and multidirectional phonological transfer (cf. Rafat 2011).

3.2.2. Final Obstruent Devoicing

The three languages under investigation differ in the realization of coda obstruents. While English retains a voicing contrast in a syllable-final position, this opposition is neutralized in German and Polish (Gonet 2001; Smith et al. 2009). Although both German and Polish manifest final obstruent devoicing, Polish additionally applies the rule of regressive voicing assimilation (Rubach 1984). Both languages have also been associated with less than a total neutralisation of the underlying voicing contrast, in that small differences in one or more acoustic properties, such as the length of the preceding vowel, have been reported when compared to underlying voiceless counterparts (Slowiaczek and Dinnsen 1985). English, in contrast to German and Polish, typically manifests the marked voiced/voiceless contrast among word final obstruents, even though individual variation has also been reported, as well as the effect of phonological environment on the production of specific word-final obstruents (Gonet 2001; Smith et al. 2009). Finally, English voiced word-final obstruents have primarily been characterized by longer duration of the preceding vowel and not necessarily by glottal pulsing (Krause 1982).

3.3. Research Questions and Hypotheses

In order to investigate cross-linguistic interactions in multilinguals' speech perception, the following research questions were posed in the study:

1. Is there evidence of CLI in the perception of L2 English and L3 German?

It is hypothesized that cross-linguistic interactions in the two foreign languages may differ and result in variable performance on the measures of perception accuracy and reaction time (RT), depending on the language status (L2 vs. L3) as well as the investigated feature (rhotics vs. final obstruent devoicing). Better performance on both measures and, thus, less CLI is expected for the more established L2 as compared to the newly acquired L3.

Hypothesis 1 (H1). *Both phonological feature and language determine perception accuracy and reaction times. There will be less CLI in the learners' L2 English than their L3 German.*

2. Is there a perceptual development over time caused by a change in CLI? Does the perceptual development in L3 parallel that in L2? In this study, cross-linguistic interactions were operationalized as the respondents' preferences for L1-, L2-, L3-accented stimuli in the performed forced-choice goodness

task. We expect different patterns to hold for the two foreign languages acquired. We expect to observe a change in CLI patterns as a function of the testing time (T1 vs. T2).

Hypothesis 2 (H2). *There will be changes in CLI across time. The developmental patterns of CLI differ between the learners' L2 English and L3 German.*

3.4. Materials and Methods

The participants performed perceptual tasks in both their L2 English and L3 German, respectively, to test their perception of rhotics and final obstruent (de)voicing. Response accuracy and reaction times were recorded for analyses at two testing times (T1, after 5 months and T2, after 10 months of L3 learning). To create appropriate language modes, the data collection for each of the languages was carried out on two separate days with L1 speakers of the respective languages as instructors.

A forced-choice (FC) goodness task was selected for the present study as an alternative to more traditional perceptual paradigms such as discrimination or identification. Perception discrimination tasks, in which the listener decides whether two stimuli are the same or different, seemed to be of little use as the aim was to test the association of a given variant of a sound with a chosen language in the multilingual's repertoire. Identification tasks in turn are inherently notorious for specifying response alternatives (including difficulties concerning non-transparent orthography), the problem being magnified in the case of three phonological systems in interaction. Moreover, identification tasks are not useful for testing allophonic differences across languages. Overly complex perception tasks needed to be avoided, too: when task complexity increases, perceivers have been found to switch to a primarily phonological level of reasoning (Strange 2009). Therefore, a forced-choice goodness task was selected for the present research, which allowed for elicitation of an association of a given allophone across multiple languages while the complexity of stimulus identification was avoided.

More specifically, the participants in this study heard two renditions of the same phrase differing on the last stimulus items embedded in a carrier phase. By pressing one of two buttons (marked 1 and 2) on a button box, they had to decide which phrase sounds more natural (i.e., more target-like) to them. One rendition was a target realization and the other was an accented language realization, where only the investigated feature was manipulated. For example, for rhotics, in the English version of the task, the stimuli included the target-like phrase "You will hear the word ring /ɹiŋ/" followed by the Polish-like realization of the rhotic sound "You will hear the word ring /riŋ/".

For rhotic sounds, this included two trials of pair items as the target item was positioned next to two other possible realizations, while for obstruent (de)voicing, it featured a single trial as the target was presented in opposition to voiced or devoiced/voiceless. The order of presentation of target and non-target stimuli was counterbalanced across trials.

Thus, in the English version, there were stimuli with English target rhotics as well as with Polish and German rhotics. Likewise, in the German version, the stimuli included German target rhotics embedded in a carrier phrase as well as Polish- and English-accented manipulated rhotics in the target words. In case of obstruent (de)voicing, the stimuli in the English version included the target-like phrase "You will hear the word have" /hæv/, followed by a manipulated realization of the final obstruent /hæf/. Similarly, in the German version, the target words embedded in a carrier phrase ("Du hörst das Wort Hand" /hant/) included final obstruents that were either voiceless (thus target like) or voiced (i.e., L2-accented).

The stimuli in each language version involved 10 pair items containing rhotics, 13 to 14 pair items featuring final (de)voicing, and three training pair items that preceded the testing blocks. In total, the FC task, thus, included 26 English and 27 German pair items for the participants to respond to.

The target rhotics occurred either in word-initial or medial position and included:

- For English: *ring, rabbit, red, round, giraffe* (with the manipulated items realized as having an L1-Polish-accented alveolar trill or an L3-German-accented uvular fricative).

- For German: *rot*, *Regen*, *Reise*, *Fahrrad*, *verloren* (with the manipulated items realized as having an L1-Polish-accented alveolar trill or an L2-English accented post-alveolar approximant).

The final obstruent (de)voicing stimuli were in coda positions and featured as follows:

- For English: *days*, *grab*, *leg*, *could*, *stab*, *big*, *skies*, *give*, *love*, *food*, *judge*, *have*, *rob* (with the manipulated items realized with voiceless final obstruents, which could be interpreted as either L1-Polish or L3-German-accented)
- For German: *Hand*, *Berg*, *Quiz*, *lieb*, *Kleid*, *Mund*, *Honig*, *Hund*, *Fahrrad*, *Kind*, *vierzig*, *brav*, *Korb*, *gelb* (with the manipulated items realized with voiced final obstruents, which could be interpreted as L2 English-accented).

The stimuli were randomized across trials in E-prime. The inter-stimulus interval was set at 500 ms and the participants had a 3000 ms response limit, thus, the task was timed. The participants' performance on the timed forced-choice goodness task was examined in terms of accuracy and reaction time (RT). The latter was included as a proxy for the perceptual difficulty of the tested stimuli.

The stimuli were recorded by three female native speakers of the respective languages, who were fluent advanced speakers of the other two languages in the triad of languages. The stimuli were produced naturalistically to avoid artificial concatenation. To ensure naturalness, several recordings of the same items were performed and validated by selecting the ones in which the performed accented manipulation sounded the most acceptable to the researchers. The process of stimulus validation was based on the perceptual assessment of each stimulus by native speakers of the respective languages. We adopted a perceptual 'category goodness' criterion, which was deemed to have the best ecological validity given the nature of the FC goodness task administered to the participants.

As far as the three speakers who produced the stimuli are concerned, their stay in a foreign country ranged from a few months to a few years. While we acknowledge the fact that their L1 production could be affected by a highly proficient knowledge of the L2/Ln, it is debatable if the prototypical monolingual rendition should be sought as the target production of the stimuli, in the light of the recent discussions on the native monolingual norm in research on multilingual acquisition (see e.g., Sorace 2020; Kroll 2020). Moreover, monolingual speakers of German, Polish, or English are increasingly impossible to find. Therefore, it was not our goal to search for a native monolingual rendition of the target items, but rather to allow for a potential variation represented by native speakers of particular languages who are multilingual speakers themselves.

4. Results

Due to violation of the assumption of normality and homogeneity of variance of the present dataset, nonparametric tests were used for between-subjects (Mann–Whitney *U*-test) and within-subjects (Wilcoxon signed-rank test) comparisons. The statistical tests were run using STATISTICA 10. The performed analyses included perceptual development over time, feature comparison, language comparison, individual variability, and CLI analysis, which will be presented in the following subsections.

4.1. Nonparametric Tests of Perception Accuracy and RT

4.1.1. Perceptual Development over Time: Perception Accuracy at T1 and T2

The performed across-time comparison did not show much development in perception accuracy for the multilingual learners. The only statistical difference between the two testing times in the performance in L2 and L3 for the two features under investigation was found for L3 German rhotics (and not in the expected direction), in which case the perception accuracy was higher at T1 than at T2 ($Z = 4.5$, $p < 0.05$) (see Table 3).

Table 3. Perception accuracy for second language (L2) and third language (L3) at testing time one (T1) and testing time two (T2).

Language	Feature	Time	N	Perception Accuracy		Wilcoxon Matched-Pairs Test	
				Mean	SD	Z	*p*
L2 English	obstr_devoicing	T1 T2	155 155	0.55 0.51	0.50 0.50	0.75	0.4509
	rhotics	T1 T2	95 95	0.92 0.92	0.27 0.28	0.24	0.8139
L3 German	obstr_devoicing	T1 T2	149 149	0.40 0.42	0.49 0.50	0.31	0.7603
	rhotics	T1 T2	89 89	0.76 0.38	0.39 0.45	4.50	* 0.0000

* $p < 0.05$.

4.1.2. Perceptual Development over Time: RT at T1 and T2

The performed Wilcoxon matched pairs signed rank test for the comparison of reaction times (RT) at two testing times (T1 vs. T2) did not show much development over time either. The only statistically significant result was found for L3 German obstruent devoicing (Z = 2.14, $p < 0.05$), with the processing time being longer at T1 than at T2 (see Table 4).

Table 4. Reaction time (RT) for L2/L3 at T1 and T2.

Language	Feature	Time	N	RT		Wilcoxon Matched-Pairs Test	
				Mean	SD	Z	*p*
L2 English	obstr_devoicing	T1 T2	155 155	644.0 581.4	437.7 397.0	1.59	0.1122
	rhotics	T1 T2	95 95	600.3 519.3	398.9 308.6	1.88	0.0596
L3 German	obstr_devoicing	T1 T2	149 149	802.4 729.3	444.0 458.9	2.14	* 0.0325
	rhotics	T1 T2	89 89	682.6 647.7	471.7 418.2	0.22	0.8222

* $p < 0.05$.

On the whole, the results did not demonstrate much development over time in perception accuracy and processing speed as measured by means of a FC task. It appears, however, that the L2 English is the more established phonological system, while L3 German is more susceptible to changes over the two testing times (i.e., a significant change in the perception accuracy of rhotics and in processing speed for obstruent devoicing). There is no consistency though in the observed developmental changes (the decrease in RT for the perception of obstruent devoicing is as expected, whereas the decrease in accuracy of rhotics perception appears counterintuitive).

4.1.3. Feature Comparison: Perception Accuracy

In the performed feature comparison, the Mann–Whitney *U*-test for perception accuracy demonstrated statistical differences in three out of four conditions: L2 English rhotics were perceived with greater accuracy than obstruent devoicing both at T1 (Z = −6.18, $p < 0.05$) and T2 (Z = −6.51, $p < 0.05$), while for L3 German the same held true at T1 (Z = −5.19, $p < 0.05$) (see Table 5).

Table 5. Comparison of perception accuracy of features at T1 and T2.

Language	Time	Feature	N	Accuracy		Mann–Whitney *U*-Test	
				Mean	SD	Z	*p*
L2 English	T1	obstr_devoicing rhotics	159 97	0.55 0.92	0.50 0.26	−6.18	* 0.0000
	T2	obstr_devoicing rhotics	165 101	0.50 0.89	0.50 0.31	−6.51	* 0.0000
L3 German	T1	obstr_devoicing rhotics	166 97	0.40 0.73	0.49 0.41	−5.19	* 0.0000
	T2	obstr_devoicing rhotics	159 94	0.44 0.40	0.50 0.46	0.52	0.6028

* $p < 0.05$.

4.1.4. Feature Comparison: RT

When the two features were compared in terms of reaction time, only one statistical difference was attested for L3 German at T1, when RT were longer for final devoicing than for rhotics (Z = 2.98, $p < 0.05$). Otherwise, the processing speed did not differ across features (see Table 6).

Table 6. Comparison of RT to features at T1 and T2.

Language	Time	Feature	N	RT		Mann–Whitney *U*-Test	
				Mean	SD	Z	*p*
L2 English	T1	obstr_devoicing rhotics	159 97	657.4 594.0	445.6 397.1	1.02	0.3100
	T2	obstr_devoicing rhotics	165 101	585.3 512.9	425.9 305.2	1.08	0.2791
L3 German	T1	obstr_devoicing rhotics	166 97	794.9 681.2	457.9 479.6	2.98	* 0.0029
	T2	obstr_devoicing rhotics	159 94	719.9 657.8	453.0 430.5	1.25	0.2130

* $p < 0.05$.

4.1.5. Language Comparison: Perception Accuracy

To compare the perception performance across languages, a Mann–Whitney *U*-test was performed, which demonstrated statistically significant differences for three out of four conditions, i.e., the perception accuracy was higher for rhotics in L2 English than in L3 German at both T1 (Z = 4.0, $p < 0.05$) and T2 (Z = 7.63, $p < 0.05$), and for obstruent devoicing at T1 (Z = 2.7, $p < 0.05$). A higher proficiency in the more established L2 was reflected in better accuracy performance in perception (see Table 7).

Table 7. Perception accuracy comparison between L2 English and L3 German.

Feature	Time	Language	N	Accuracy		Mann-Whitney *U* Test	
				Mean	SD	Z	*p*
obstr_devoicing	T1	L2 English L3 German	159 166	0.55 0.40	0.50 0.49	2.70	* 0.0070
	T2	L2 English L3 German	165 159	0.50 0.44	0.50 0.50	1.02	0.3075
rhotics	T1	L2 English L3 German	97 97	0.92 0.73	0.26 0.41	4.00	* 0.0001
	T2	L2 English L3 German	101 94	0.89 0.40	0.31 0.46	7.63	* 0.0000

* $p < 0.05$.

4.1.6. Language Comparison: RT

A Mann–Whitney *U*-test for reaction time comparison between L2 English and L3 German demonstrated statistically significant differences for three out of four conditions, i.e., RTs were longer in L3 German than in L2 English for the perception of obstruent devoicing at both T1 and T2 and for the perception of rhotics at T2. On the whole, it took longer to process the perception task in the L3 than in the L2 (see Table 8).

Table 8. Reaction time comparison between L2 English and L3 German.

Feature	Time	Language	N	RT		Mann–Whitney *U*-Test	
				Mean	SD	Z	*p*
obstr_devoicing	T1	L2 English L3 German	159 166	657.4 794.9	445.6 457.9	−3.70	* 0.0002
	T2	L2 English L3 German	165 159	585.3 719.9	425.9 453.0	−3.26	* 0.0011
rhotics	T1	L2 English L3 German	97 97	594.0 681.2	397.1 479.6	−0.74	0.4606
	T2	L2 English L3 German	101 94	512.9 657.8	305.2 430.5	−2.36	* 0.0184

* $p < 0.05$.

4.1.7. Correlation: Perception Accuracy and RT

No statistically significant correlations were found between perception accuracy and reaction time for L2 English and L3 German performance in the perception of rhotics and final devoicing at either T1 or T2 (see Table 9).

Table 9. Correlation between perception accuracy and RT at T1 and T2.

Language	Time	Feature	n	r(X.Y)	t	*p*
L2 English	T1	obstr_devoicing rhotics	13 8	0.371 −0.618	1.32 −1.92	0.2121 0.1027
	T2	obstr_devoicing rhotics	13 8	0.288 −0.577	1.00 −1.73	0.3400 0.1345
L3 German	T1	obstr_devoicing rhotics	14 8	0.203 0.287	0.72 0.73	0.4859 0.4911
	T2	obstr_devoicing rhotics	14 8	0.062 −0.099	0.22 −0.24	0.8333 0.8160

4.2. GLM Modelling

We fitted our data to a generalized linear model (GLM), with the dependent variable being perception accuracy and independent variables including RT, testing time (T1 and T2) and feature (obstruent devoicing and rhotics). The analysis was performed separately for each language and based on the number of token items rather than participants.

The GLM analysis for L2 English revealed a significant effect of feature on the perceptual accuracy in L2 English [$F(1,522) = 92.79$, $p < 0.05$)], while the testing time and RT were not significant predictors (see Table 10).

Table 10. Results of a linear model for the dependent variable—Accuracy for L2 English.

Effect	SS	df	MS	F	*p*
Intercept	94.55	1.00	94.55	506.05	* 0.0000
RT	0.42	1.00	0.42	2.27	0.1326
Testing time	0.30	1.00	0.30	1.59	0.2075
Feature	17.34	1.00	17.34	92.79	* 0.0000
Time*Feature	0.02	1.00	0.02	0.10	0.7559
Error	96.59	517.00	0.19		

* $p < 0.05$.

The Bonferroni pairwise comparisons confirmed that there were statistically significant differences ($p < 0.001$) between perception accuracy for rhotics and obstruent devoicing, with the former feature generating higher accuracy rates (see Table 11).

Table 11. Mean Accuracy with respect to Feature for L2 English.

Feature	n	Mean	SD	SE
obstr_devoicing	324	0.52	0.50	0.03
rhotics	198	0.91	0.29	0.02

The GLM analysis for L3 German failed to find a significant effect of RT, however, the remaining variables proved to be significant predictors for perceptual accuracy in L3 German, namely testing time [($F(1, 516) = 11.85$, $p = 0.000$)], feature [($F(1, 516) = 10.55$, $p = 0.001$)], and the Time*Feature interaction [($F(1, 516) = 18.05$, $p = 0.000$)] (see Table 12).

Table 12. Results of a linear model for the dependent variable—Accuracy for L3 German.

Effect	SS	df	MS	F	*p*
Intercept	38.07	1.00	38.07	168.58	* 0.0000
RT	0.06	1.00	0.06	0.28	0.5938
Testing Time	2.68	1.00	2.68	11.85	* 0.0006
Feature	2.38	1.00	2.38	10.55	* 0.0012
Time*Feature	4.08	1.00	4.08	18.05	* 0.0000
Error	115.40	511.00	0.23		

* $p < 0.05$.

The Bonferroni pairwise comparisons pointed to a statistically significant difference ($p = 0.017$) between the two testing times in L3 German, with higher perception accuracy observed at T1 (see Table 13).

Table 13. Mean Accuracy with respect to Testing Time for L3 German.

Testing Time	N	Mean	SD	SE
T1	263	0.52	0.49	0.03
T2	253	0.42	0.48	0.03

Bonferroni correction confirmed a statistically significant difference between perceptual accuracy of the two investigated features (p = 0.0008), with rhotics being perceived more accurately than obstruent devoicing in L3 German (Table 14).

Table 14. Mean Accuracy with respect to Feature for L3 German.

Feature	N	Mean	SD	SE
obstr_devoicing	325	0.42	0.49	0.03
rhotics	191	0.57	0.47	0.03

The Bonferroni pairwise comparisons confirmed that there were statistically significant differences for perceptual accuracy in L3 German between the following variables: (1) obstruent devoicing at T1 and rhotics at T1 ($p < 0.0001$); (2) obstruent devoicing at T2 and rhotics at T1 ($p < 0.0001$); (3) rhotics at T1 and rhotics at T2 ($p < 0.0001$) (see Table 15).

Table 15. Mean Accuracy with respect to the Time*Feature interaction for L3 German.

Testing Time	Feature	N	Mean	SD	SE
T1	obstr_devoicing	166	0.40	0.49	0.04
T1	rhotics	97	0.73	0.41	0.04
T2	obstr_devoicing	159	0.44	0.50	0.04
T2	rhotics	94	0.40	0.46	0.05

4.3. Individual Differences

Figures 1–8 show that, in general, more inter- and intraspeaker variability occurs in L3 German than in L2 English, in which individual perceptual performance seems more homogeneous across learners. This is especially true for the perception of the English rhotic where six learners show ceiling performance at both testing times. Pronounced changes in perception accuracy across time are, however, apparent for individual learners. In the case of Subject 20, for instance, their perception of both L2 English obstruent voicing and rhotics drops drastically from T1 to T2 and also shows a drop in perception accuracy in the L3 German rhotic from well above chance to well below it from T1 to T2 (see Figures 1–4, 7 and 8). Subject 12, in turn, performs consistently accurately in their perception of the L2 sounds under examination (Figures 1–4). Their perception of the L3 counterparts drops between the two testing times, most dramatically for rhotics (Figures 5–8). Some increase in L2 English perception of final obstruents (Figures 1 and 2) together with a dramatic improvement of L3 German perception of the same feature (Figures 5 and 6) was evidenced in Subject 6. See Figures 1–8, illustrating perception accuracy of individual subjects in L2 English and L3 German at T1 and T2 for obstruent devoicing and rhotics (with group means marked as horizontal black lines on the graphs).

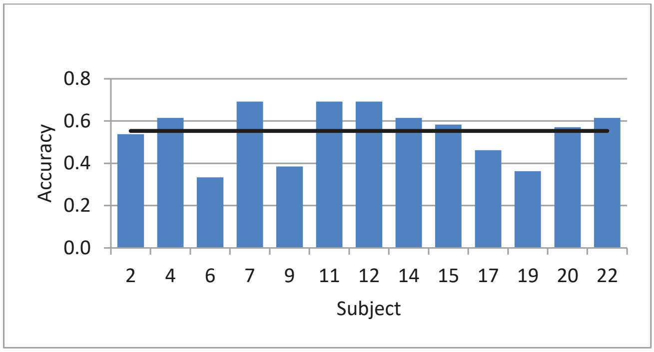


Figure 1. Perception accuracy in L2 English for obstruent devoicing at T1.

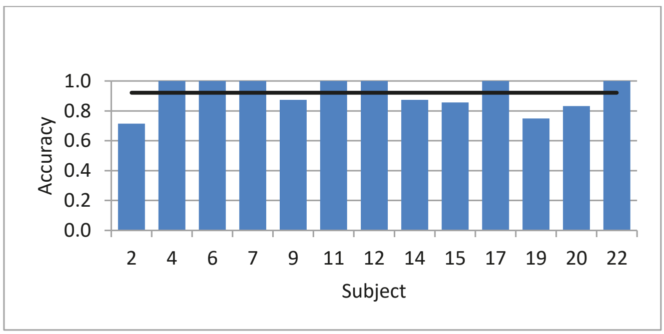


Figure 2. Perception accuracy in L2 English for rhotics at T1.

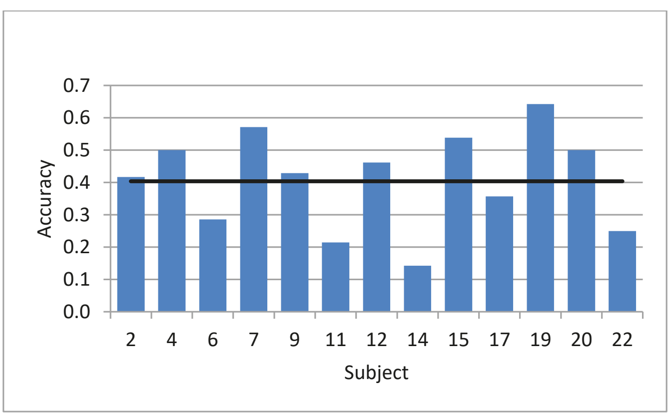


Figure 3. Perception accuracy in L3 German for obstruent devoicing at T1.

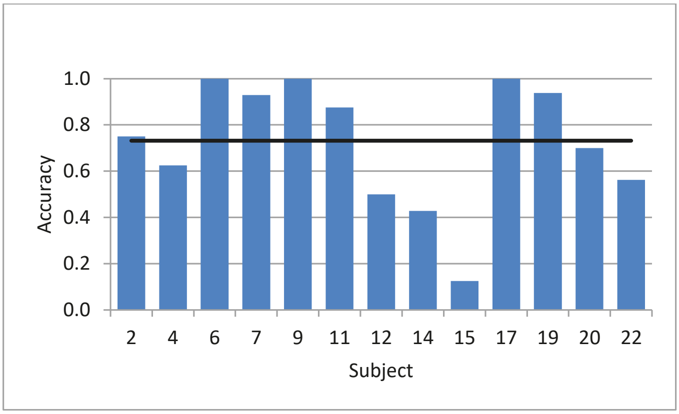


Figure 4. Perception accuracy in L3 German for rhotics at T1.

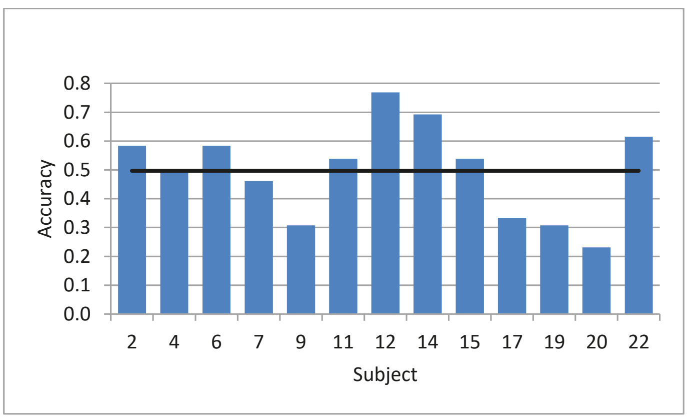


Figure 5. Perception accuracy in L2 English for obstruent devoicing at T2.

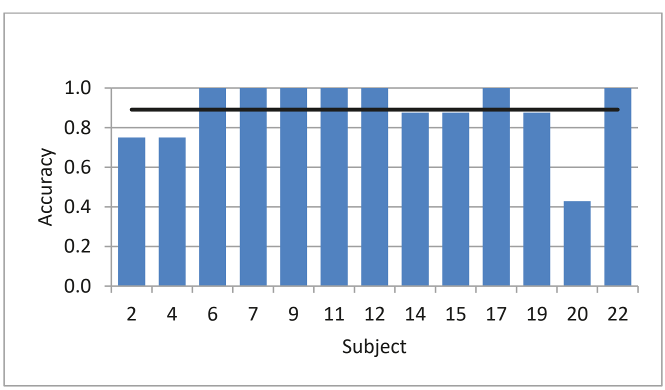


Figure 6. Perception accuracy in L2 English for rhotics at T2.

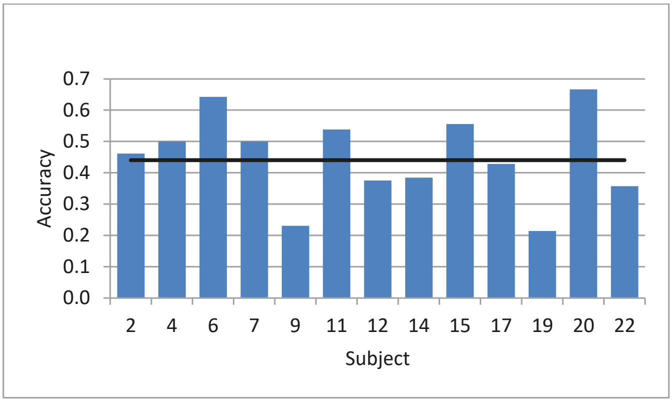


Figure 7. Perception accuracy in L3 German for obstruent devoicing at T2.

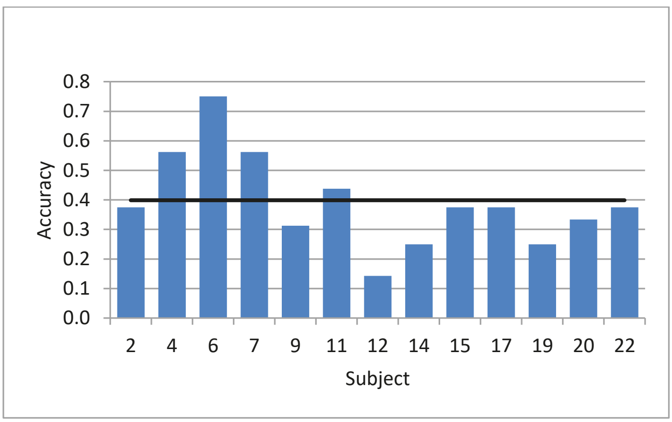


Figure 8. Perception accuracy in L3 German for rhotics at T2.

4.4. CLI

In order to explore cross-linguistic mappings in the perception of the multilingual learners of this study, we further explored their perception accuracy (as the dependent variable) with respect to the different stimulus properties of the perception task employed (i.e., L1-accented, L2-accented, L3-accented) as independent variables.

For rhotics, the performed ANOVA (with L2 and L3 treated jointly) demonstrated that there was a statistically significant difference in perception accuracy between these three conditions ($F(2;24) = 46.38$, $p < 0.05$). The post-hoc Scheffé test for multiple comparisons showed that the differences between all pairs of differently accented stimuli were significant ($p < 0.05$). The accuracy of perceiving the correct rhotic stimuli in L2 English was the highest when the other manipulated stimulus was L3-(German) accented, while it was the least accurate when the unnatural stimulus was L2-(English) accented in L3 German (see Figure 9). Interestingly, however, when we compared the latencies of responses in all these conditions, there were no statistical differences found in RT for rhotics irrespective of the source of accent in the manipulated stimuli.

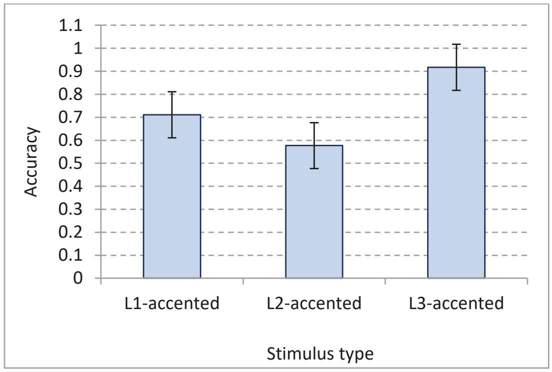


Figure 9. Perception accuracy of rhotics according to stimuli types (L1-accented, L2-accented, L3-accented), with 0.95 confidence intervals as whisker bars.

For final devoicing, given the binary response option as well as difficulty in strictly disentangling L1-based source of CLI in the perception of this feature from arguably the lack of it (L3-target stimuli), the results evidence CLI primarily from L1 and/or L3 in the case of perceiving L2 final obstruent voicing (accuracy levels at chance levels, with acceptance of L1/L3-based and L2-based stimuli to comparable levels). However, in the case of L3 final obstruent devoicing, L1-based CLI prevailed (L1-accented/L3-based stimuli were generally perceived as being more natural than L2-accented stimuli) (t = 4.12, $p < 0.05$).

As far as the reaction time is concerned, none of the independent variables (i.e., feature, stimulus type) entered into the GLM analysis proved to be significant, nor did the interaction between feature and stimulus type ($p > 0.05$). It follows that no statistical differences were found in RT, irrespective of the source of accent in the manipulated stimuli, in the perception of both of the investigated features, although there was a visible trend for the L2-accented stimuli in the perception of L3 obstruent devoicing taking longer to process that the L2-accented stimuli in L3 rhotics.

5. Discussion

Our results show that the effects of CLI on multilinguals' perception differ across both their two languages and the two features under investigation, thus confirming Hypothesis 1. Overall, perception accuracy is higher in their L2 English than in their L3 German and processing speed is faster, as predicted by Hypothesis 1. Moreover, perception accuracy in the L2 English, which they had been learning for 5–6 years, is more stable across time than for the L3 German, confirming Hypothesis 2. Our results, thus, suggest that CLI is lowest for the perception of the L2 English, especially for rhotics, where most of the investigated learners seem to have established stable perceptual categories. However, we did not test learners' perception in their L2 English after a few weeks of learning the new language German, and, thus, might have missed the short-term effect of influence from the new L3, the 'perceptual confusion' found by Nelson (2020). In fact, one individual learner did show a drop in L2 perception accuracy even after ten months of learning the L3, which might have the same underlying cause.

Our results further show that overall perception accuracy is higher for rhotics than for final obstruent in both languages. Perception of final obstruent devoicing in both the learners' L2 and L3 is at chance level at both testing times, evidencing no improvement for any of the learners, while perception of the rhotics is significantly higher in both languages, with individual speakers reaching ceiling performance. Contrary to the predictions of our Hypothesis 1, this suggests a high level of CLI for the former feature, even in the L2 English, for which learners had been attending school lessons for 5–6 years. One explanation might be the lower perceptual saliency of final obstruent (de)voicing compared to the different articulations of the rhotics in the three languages under investigation. Moreover, the phonological process of obstruent voicing in coda position is characterized by a complex interaction of phonetic cues beyond that of glottal pulsing (Krause 1982). As shown in Broesma (2005), even highly proficient learners of English do not use native-like weighing of cues for the perception of voicedness in an unfamiliar position. Our learners may, thus, have had a hard time to attend to the relevant phonetic cues, longer duration for the preceding vowel in particular, to distinguish between the pairs of tested stimuli.

By the same token, evidence for CLI was found in the learners' perception in their L3 German: their accuracy of perceiving the German rhotic /R/ was higher after 5 months of learning than after 10 months. It appears that some restructuring of perceptual categories is still under way in the first ten months of exposure to a new language, thus echoing findings by Balas et al. (2019). However, again, this restructuring seems to be feature-dependent rather than a general mechanism as these changes were found only for the perception of the rhotics but not for the perception of final obstruent devoicing. Our findings thus appear to partially contradict the predictions of PAM-L2 (Best and Tyler 2007), which would expect a continuous refinement of the learners' speech perception as a function of their extended experience with learning the language. Possibly, this refinement only takes place after more input than our learners had enjoyed in their L3 after 10 months of learning. Not incompatible with this line of reasoning, it might be that the L3 learners in this study had been increasingly exposed to foreign-accented realizations of the German rhotic sound in their classroom environment, whether from their peers or their Polish teacher of German, thus, developing a nontarget representation of naturalness for it. Their own experience with producing the articulatorily challenging sound in the first year of learning German may have also contributed to the process of their category formation for the sound (cf. Bundgaard-Nielsen et al. 2011). As an alternative explanation for the drop in perceptual performance, one could point to a possible decreased attention to the task at the second testing point as compared with the novelty of the first testing time that triggered more focused interest and auditory processing in the participants. This finding would be in line with Nelson (2020) observations concerning the initial 'novelty effect' in perceptual performance of her child and adult L3 learners.

The source of cross-linguistic influence on the perception accuracy of the multilingual learners was found to vary: the accuracy of perceiving the L2 English rhotic /ɹ/ was higher when it was contrasted with an L3-German accented stimulus than with an L1-Polish accented stimulus in the FC task. This would point towards a stronger influence of the L1 than the L3 in the perception of rhotics in the L2, although recall that overall L2 rhotics were perceived highly accurately by the learners. On the other hand, in L3 German, the accuracy of perceiving the rhotic was lowest when it was contrasted with an L2 English accented stimulus rather than with an L1-Polish stimulus, which leads to the conclusion that the L2 rather than the L1 was a stronger source of CLI for the L3 perception of this feature. This would seem to suggest initially a greater influence of the L2 than the L1 on the perception of L3 rhotics, a finding that was also reported in Wrembel et al. (2019). Indeed, initial L3 learners appear to map new non-native phones to both their L1 and L2, which may be interpreted as aligning with the general reasoning of most L2 speech perception models: non-native phones are perceived in relation to previously established (or currently being established) categories depending on the degree of perceived cross-linguistic similarity between the phones concerned. The way in which such perceived cross-linguistic mappings are to be most effectively elicited in multilingual perceivers presents one of the greatest methodological challenges in future L3 speech research.

Regarding obstruent devoicing, it was not possible to disentangle the sources of CLI for L2 perception (due to the identical nature of L1-accented and L3-based stimuli). However, if we assume the existence of CLI at this stage of L3 learning, L3 perception of the devoicing feature was arguably influenced more significantly by the L1 than the L2, considering the more marked status of obstruent voicing in L2 as well as the similar standing of this feature in the L1 and L3.

Our results further showed that factors other than CLI might influence speech perception. Higher accuracy in the L2 than in the L3 and the fact that L2 is processed faster than the L3 are viewed as evidence that what also matters in non-native speech perception is experience. Our results corroborate the effects of language learning experience on non-native consonant perception, similarly to some previous studies (e.g., Bradlow et al. 1999; Rose 2010; Anderson 2011), which reported some improvement for more experienced participants or after perception training, but also considerable variation across subjects and the phones tested, as predicted by L2LPM (Escudero and Boersma 2004; Escudero 2005, 2009).

No correlation was found between the learners' perception accuracy and reaction time in the perception of rhotics and final devoicing in either language and at either observation point. This suggests that processing speed is quite independent of the degree of establishment of perceptual categories and may not be the most informative proxy for evaluations of the learnability of different sounds, at least for L3 learning contexts.

As for the role of markedness, in the present study we tested one feature which was more marked in the L2 than in the L1 and L3 (i.e., final obstruent devoicing) and one feature which was more marked in the L3 than in the L2 (i.e., German uvular vs. English postalveolar rhotics). L2 English rhotics were more accurately chosen when contrasted with L3 German stimuli, possibly suggesting a stronger influence of the less marked L1 rhotic than of the most marked L3 rhotic on the L2 perception of a relatively unmarked rhotic variant. Contrastively, in L3 German, the less marked L2 rhotic influenced perception to a greater extent than the more marked L1 rhotic. In final obstruent devoicing, the accuracy was around or below the chance level, and it seems the more marked L2 variant has not been internalized by the learners at all. Therefore, in order to further disentangle the influence of language status from markedness of the tested feature, more studies that would use various combinations of markedness and language status are needed.

6. Conclusions

The overall results indicate that CLI in perceptual development is feature-dependent with relative stability evidenced for L2 rhotics, reverse trends for L3 rhotics, and no significant development for L2/L3 (de)voicing. We also found that perception accuracy of rhotics differed significantly with respect to stimulus properties, (i.e., whether they were L1-accented, L2-accented, or L3-accented) and that it took longer to process the perception task in the L3 than L2. On the whole, major findings include a nonlinear development of foreign language phonology, diverse CLI patterns that are feature-dependent, and differential learnability of phonetic features. We hope the present findings will be an incentive to extend current theoretical frameworks beyond L2 speech perception models to account for these phenomena in multilingual speech perception.

Author Contributions: Conceptualization, M.W., R.K., A.B., and U.G.; methodology, M.W., R.K., A.B. and U.G.; formal analysis, M.W. and R.K.; investigation, M.W., R.K., A.B., and U.G.; data curation, M.W., R.K. and A.B.; writing—original draft preparation, M.W., R.K., A.B., and U.G.; writing—review and editing, M.W., R.K., A.B., and U.G.; visualization, M.W.; project administration, U.G.; funding acquisition, M.W., R.K., A.B., and U.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Polish-German Foundation of Science [project no. 2017-10].

Acknowledgments: We wish to acknowledge the assistance of Iga Krzysik and Halina Lewandowska in the data collection process.

Conflicts of Interest: The authors declare no conflict of interest.

References

Anderson, Gregory. 2011. Non-Linear Dynamics of Adult Non-Native Phoneme Acquisition Perception & Production. Ph.D. thesis, George Mason University, Fairfax, VA, USA.

Baker, Wendy, Pavel Trofimowich, Molly Mack, and James E. Flege. 2002. The effect of perceived phonetic similarity on non-native sound learning by children and adults. In *Proceedings of the 26th Annual Boston University Conference on Language Development*. Edited by Barbora Skarabela, Sarah Fish and Anna H. -J. Do. Somerville: Cascadilla, pp. 36–47.

Balas, Anna, Romana Kopečková, and Magdalena Wrembel. 2019. Perception of rhotics by multilingual children. In *Proceedings of the 19th International Congress of Phonetic Sciences*. Edited by Sasha Calhoun, Paola Escudero, Marija Tabain and Paul Warren. Melbourne: Australasian Speech Science and Technology Association, pp. 3725–29.

Best, Catherine T. 1995. A direct realist view of cross-language speech perception. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Edited by Winifred Strange. Baltimore: York Press, pp. 171–204.

Best, Catherine T., and Michael D. Tyler. 2007. Non-native and second language speech perception: Commonalities and complementaries. In *Second Language Speech Learning*. Edited by Murray J. Munro and Ocke-Schwen Bohn. Amsterdam: John Benjamins Publishing, pp. 13–34.

Bradlow, Ann. R., Reiko Akahane-Yamada, David B. Pisoni, and Yoh'ichi Tohkura. 1999. Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception and Psychophysics* 61: 977–85. [CrossRef] [PubMed]

Broesma, Mirjam. 2005. Perception of familiar contrasts in unfamiliar positions. *Journal of the Acoustical Society of America* 117: 3890–901. [CrossRef] [PubMed]

Bundgaard-Nielsen, Rikke. L., Catherine T. Best, and Michael D. Tyler. 2011. Vocabulary size matters. The assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Applied Psycholinguistics* 32: 51–67. [CrossRef]

Cabrelli, Amaro Jennifer, and Magdalena Wrembel. 2016. Investigating the acquisition of phonology in a third language—A state of the science and an outlook for the future. *International Journal of Multilingualism* 13: 395–409. [CrossRef]

Cebrian, Juli. 2006. Experience and the use of no-native duration in L2 vowel categorization. *Journal of Phonetics* 34: 372–87. [CrossRef]

Cenoz, Jasone, and Ulrike Jessner. 2000. *English in Europe: The Acquisition of a Third Language*. Clevedon: Multilingual Matters.

Cenoz, Jasone, Britta Hufeisen, and Ulrike Jessner. 2001. *Cross-Linguistic Influence in Third Language Acquisition: Psycholinguistic Perspectives*. Clevedon: Multilingual Matters.

De Angelis, Gessica. 2007. *Third or Additional Language Acquisition*. Clevedon: Multilingual Matters.

Eckman, Fred. 1977. Markedness and the contrastive analysis hypothesis. *Language Learning* 27: 315–30. [CrossRef]

Escudero, Paola. 2005. Linguistic Perception and Second Language Acquisition. LOT Dissertation Series 113. Ph.D. thesis, Utrecht University, Utrecht, The Netherlands.

Escudero, Paola. 2009. Linguistic perception of 'similar' L2 sounds. In *Phonology in Perception*. Edited by Paul Boersma and Silke Hamann. Berlin: Mouton de Gruyter, pp. 151–90.

Escudero, Paola, and Paul Boersma. 2004. Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26: 551–85. [CrossRef]

Flege, James E. 1991. Orthographic evidence for the perceptual identification of vowels in Spanish and English. *The Quarterly Journal of Experimental Psychology* 43A: 701–31. [CrossRef] [PubMed]

Flege, James E. 1995. Second language speech learning: Theory, problems, findings. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Edited by Winifred Strange. Baltimore: York Press, pp. 233–77.

Flege, James E., and Ocke-Schwen Bohn. 2020. The revised Speech Learning Model. Available online: https://www.researchgate.net/publication/342923320_The_revised_Speech_Learning_Model (accessed on 19 August 2020).

Flynn, Suzanne, Clair Foley, and Inna Vinnitskaya. 2004. The Cumulative-Enhancement Model for language acquisition: Comparing adults' and children's patterns of development in L1, L2 and L3 acquisition of relative clauses. *The International Journal of Multilingualism* 1: 3–16. [CrossRef]

Gonet, Wiktor. 2001. Voicing control in English and Polish: A pedagogical perspective. *International Journal of English Studies* 1: 73–92.

Gut, Ulrike. 2010. Cross-linguistic influence in L3 phonological acquisition. *International Journal of Multilingualism* 7: 19–38. [CrossRef]

Gut, Ulrike, Robert Fuchs, and Eva-Maria Wunder. 2015. *Universal or Diverse Paths to English Phonology*. Berlin: De Gruyter Mouton.

Hufeisen, Britta, and Beate Lindemann. 1997. *Tertiärschprachen, Theorien, Modelle, Methoden*. Tübingen: Staufenburg.

Jassem, Wiktor. 2003. Polish. *Journal of the International Phonetic Association* 33: 103–7. [CrossRef]

Kohler, Klaus. 1999. German. In *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press, pp. 86–89.

Kopečková, Romana. 2012. Differences in L2 segmental perception: The effects of age and L2 experience. In *Intensive Exposure Experiences in Second Language Learning*. Edited by Carmen Muñoz. Clevedon: Multilingual Matters, pp. 234–55.

Krause, Sue Ellen. 1982. Developmental use of vowel duration as a cue to postvocalic stop consonant voicing. *Journal of Speech, Language and Hearing Research* 25: 388–93. [CrossRef] [PubMed]

Kroll, Judith. 2020. Learning and Using Two Languages May Change Your Mind. Available online: http://education.uci.edu/uploads/7/2/7/6/72769947/bb_kroll_20200420.pdf (accessed on 3 November 2020).

Ladefoged, Peter. 2001. *Vowels and Consonants: An Introduction to the Sounds of Languages*. Oxford: Blackwells.

Ladefoged, Peter, and Ian Maddieson. 1996. *The Sounds of the Worlds' Languages*. Oxford: Blackwell.

Lado, Robert. 1957. *Linguistics across Cultures: Applied Linguistics for Language Teachers*. Ann Arbor: University of Michigan Press.

Nelson, Christina. 2020. The younger, the better? Speech perception development in adolescent vs. adult L3 learners. In *Young Linguists' Insights: From Exploration to Explanation in Psycho- and Sociolinguistic Eesearch*. Edited by Katarzyna Jankowiak and Joanna Pawelczyk. Special Issue, Yearbook of the Poznań Linguistic Meeting. vol. 6, pp. 27–58.

Rafat, Yasaman. 2011. Orthography-Induced Transfer in the Production of Novice Adult English-speaking Learners of Spanish. Ph.D. thesis, University of Toronto, Toronto, ON, Canada.

Rallo Fabra, L., and Gallero J. Romero. 2012. Native Catalan learners' perception and production of English vowels. *Journal of Phonetics* 40: 491–508. [CrossRef]

Rose, Marda. 2010. Differences in discriminating L2 consonants: A comparison of Spanish taps and trills. In *Selected Proceedings of the 2008 Second Language Research Forum*. Edited by Matthew T. Prior, Yukiko Watanabe and Sang-Ki Lee. Somerville: Cascadilla Proceedings Project, pp. 181–96.

Rubach, Jerzy. 1984. *Cyclic and Lexical Phonology: The Structure of Polish*. Dordrecht: Foris.

Slowiaczek, Louisa M., and Daniel A. Dinnsen. 1985. On the neutralizing status of Polish word-final devoicing. *Journal of Phonetics* 13: 325–41. [CrossRef]

Smith, Bruce L., Rachel Hayes-Harb, Michael Bruss, and Amy Harker. 2009. Production and perception of voicing and devoicing in similar German and English word pairs by native speakers of German. *Journal of Phonetics* 37: 257–75. [CrossRef]

Sorace, Antonella. 2020. L1 attrition in a wider perspective. *Second Language Research*. [CrossRef]

Strange, Winifred. 2009. Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics* 39: 456–66. [CrossRef]

Tyler, Michael. 2019. PAM-L2 and phonological category acquisition in the foreign language classroom. In *A Sound Approach to Language Matters—In Honor of Ocke-Schwen Bohn*. Edited by Anne Mette Nyvad, Michaela Hejná, Anders Højen, Anna Bothe Jespersen and Mette Hjortshøj Sørensen. Denmark: Department of English, School of Communication & Culture, Aarhus University, pp. 607–30.

Wrembel, Magdalena. 2015. *In Search of a New Perspective: Cross-Linguistic Influence in the Acquisition of Third Language Phonology*. Poznań: Wydawnictwo Naukowe UAM.

Wrembel, Magdalena, and Jennifer Amaro Cabrelli, eds. 2018. *Advances in the Investigation of L3 Phonological Acquisition*. London: Routledge.

Wrembel, Magdalena, Marta Marecka, and Romana Kopečková. 2019. Extending perceptual assimilation model to L3 phonological acquisition. *International Journal of Multilingualism.* 16: 513–33. [CrossRef]

Wunder, Eva-Maria. 2014. On the Hunt for Lateral Phonological Crosslinguistic Influence in Third or Additional Language Acquisition. Ph.D. thesis, University of Münster, Münster, Germany.

languages

Article

Short-Term Sources of Cross-Linguistic Phonetic Influence: Examining the Role of Linguistic Environment

Daniel J. Olson

School of Languages and Cultures, Purdue University, West Lafayette, IN 47907, USA; danielolson@purdue.edu

Received: 16 August 2020; Accepted: 13 October 2020; Published: 24 October 2020

Abstract: While previous research has shown that bilinguals are able to effectively maintain two sets of phonetic norms, these two phonetic systems experience varying degrees of cross-linguistic influence, driven by both long-term (e.g., proficiency, immersion) and short-term (e.g., bilingual language contexts, code-switching, sociolinguistic) factors. This study examines the potential for linguistic environment, or the language norms of the broader community in which an interaction takes place, to serve as a source of short-term cross-linguistic phonetic influence. To investigate the role of linguistic environment, late bilinguals (L1 English—L2 Spanish) produced Spanish utterances in two sessions that differed in their linguistic environments: an English-dominant linguistic environment (Indiana, USA) and a Spanish-dominant linguistic environment (Madrid, Spain). Productions were analyzed at the fine-grained acoustic level, through an acoustic analysis of voice onset time, as well as more holistically through native speaker global accent ratings. Results showed that linguistic environment did not significantly impact either measure of phonetic production, regardless of a speaker's second language proficiency. These results, in conjunction with previous results on long- and short-term sources of phonetic influence, suggest a possible primacy of the immediate context of an interaction, rather than broader community norms, in determining language mode and cross-linguistic influence.

Keywords: bilingualism; phonetics; language mode; cross-linguistic influence; transfer; voice onset time; global accent rating

1. Introduction

Research has shown that bilinguals, including both early bilinguals (e.g., MacLeod and Stoel-Gammon 2005) and late second language learners (e.g., Schmid et al. 2014), can effectively maintain two separate phonetic systems for their two languages. However, these two phonetic systems are not fully independent, and cross-linguistic influence, in which the phonetic system of one language is influenced by the competing language, has been evidenced across a range of bilingual populations and contexts. Importantly, there are a variety of both long-term and short-term sources of cross-linguistic influence (i.e., transfer), impacting both a bilingual's first (L1) and second (L2) languages. Broadly, short-term refers to contexts in which production or perception may be altered for a single speaker in response to immediate or momentary changes in the linguistic situation (e.g., bilingual language mode and code-switching), while long-term refers to sustained influences over longer periods of times (e.g., acquisition and immersion). While some long-term sources of phonetic cross-linguistic influence, such as immersion (e.g., Casillas 2020) and instruction (e.g., Lee et al. 2015), are well-studied, less research has focused on potential short-term (i.e., transient) sources of cross-linguistic influence (Simonet 2014).

Given the previous focus on long-term sources of cross-linguistic phonetic influence, and the emerging research showing the relevance of a number of short-term sources, the current study examines the potential impact of a novel source of such influence: linguistic environment. Linguistic environment is broadly defined as the language norms of the broader community in which an interaction or experimental paradigm is conducted. Bilinguals naturally move from one linguistic environment to another for work, travel, and social interaction, and such shifts in context or environment may serve to foster cross-linguistic influence at the phonetic level. This study adds to our theoretical understanding of the organization of bilingual phonetic systems, and highlights both the sources of and limits on cross-linguistic phonetic influence.

2. Literature Review

Cross-linguistic influence at the phonetic level can be described as the way in which the phonetic system of one of a bilingual's languages impacts the production and perception of speech sounds in their other language (Jarvis and Pavlenko 2008). Within studies of bilingual phonetics, and as a condition for examining cross-linguistic influence or transfer, early research sought to establish that bilinguals are indeed able to produce and maintain separate phonetic norms in each of their two languages (e.g., Caramazza et al. 1973). While bilinguals produce different phonetic categories for their two languages, the relationship of these categories to the monolingual norms may depend on a variety of factors. For example, some research has shown that bilinguals, particularly early bilinguals, may show little to no deviance from the monolingual targets in each of their languages (e.g., Flege et al. 1999; Guion et al. 2004; Piske et al. 2002), while others have found that late bilinguals (e.g., Flege and Eefting 1987; Flege and Port 1981) and even some early bilinguals (e.g., Flege et al. 1995; Fowler et al. 2008) produce phonetic categories that deviate from those produced by monolingual speakers.

2.1. Long-Term Sources of Cross-Linguistic Phonetic Influence

In the line of research that has examined cross-linguistic phonetic influence, there has been a significant body of research that has examined the impact of L1 phonetic systems on L2 phonetic categories. Broadly, this line of research has established that the extant L1 system exerts influence over the L2 system, shaping both production and perception of L2 phonetics. Several L2 phonological models provide theoretical accounts for the mechanisms that govern the acquisition of new phonetic categories, and there is broad agreement that the ability to acquire a new L2 category depends on the relationship to existing L1 sounds (e.g., Speech Learning Model (Flege 1987, 1988, 1991, 1995); Native Language Magnet theory (Kuhl 1992, 1993a, 1993b); Perceptual Assimilation Model-L2 (Best and Tyler 2007)). Moreover, as the L2 phonetic system develops as the result of engagement with the L2, either through immersion (for review of study abroad and at-home instruction, see Casillas (2020); for longer-term immigration, see Piske et al. (2001); for study abroad, see Solon and Long (2018))[1] or instructed acquisition (for review of L2 phonetic instruction, see Lee et al. (2015)), the cross-linguistic influence of the L1 on the L2 diminishes and the L2 becomes more native-like. As shown in research on study abroad, such effects may be observed following stays of just a few weeks (e.g., Lord 2010), although Solon and Long (2018) note significant individual variation. Following long-term exposure, ultimate phonetic attainment may (e.g., Schmid et al. 2014) or may not (e.g., Flege 1987, 1991) match native speaker norms (for variability in attainment, see Simon (2009)).

Beyond the expected influence of long-term exposure to the L2 on L2 phonetic production, there is also evidence of influence of the L2 on the L1 phonetic system. Several studies have found an inverse relationship between the length of residence in an L2 linguistic environment or L2 proficiency and the degree of phonetic influence of the L2 on the L1 (e.g., Bergmann et al. 2016; Major 1992;

[1] For a discussion on the interaction between length of residence and other variables, notably age of acquisition, see Piske et al. (2001). As Piske et al. (2001) note, length of residence "only provides a rough index of overall L2 experience" (p. 197). While many studies talk about immersion, immigration, and length of residence, these may be used as a proxy for overall L2 experience. For the current study, it is relevant that these factors, whether conceptualized as L2 experience or length of residence, function as long-term sources of linguistic change and interaction.

Stoehr et al. 2017). Speakers that had spent longer in the L2 linguistic environment, or those with greater proficiency in the L2, evidenced greater degrees of L2 phonetic transfer to the L1 (e.g., Major 1992), leading Major (1992) to claim that the L1 phonetic system "is not a fixed and stable system but rather a fluid and changeable one that is highly subject to the influence of a well-developed second system" (Major 1992, p. 204). Other research has shown a degree of bidirectional cross-linguistic influence (e.g., Fowler et al. 2008), in which the L1 influences the L2 and the L2 influences the L1. Again, cases of L2 to L1 transfer have most often been found following long-term engagement with the L2.

Although much of this research has focused on a linear trajectory, such as tracking the shift in L2 production towards monolingual norms over time, it is clear that such long-term shifts are dynamic. In their seminal work, Sancier and Fowler (1997) tracked the phonetic production of a single Portuguese-English bilingual speaker as they moved between multiple linguistic environments. Their results showed that, following a stay of several months in a Portuguese linguistic environment, voice onset times (VOTs) for both Portuguese and English productions became shorter and more Portuguese-like, a phenomenon referred to as "gestural drift" (Sancier and Fowler 1997, p. 422). In contrast, following a stay of several months in an English linguistic environment, VOTs for both languages became longer and more English-like. These results demonstrate that a long-term change in linguistic environment can promote a degree of phonetic interaction, whereby both of the bilingual's languages are impacted by the language of the ambient environment. In short, cross-linguistic phonetic influence, as evidenced by bilingual gestural drift, appears to be dynamic and responds long-term to factors in the ambient linguistic environment.

The existing studies on long-term immersion, study abroad, and gestural drift generally consider changes in L1 and L2 phonetic production following a significant length of stay in the L2 linguistic environment, from weeks or months (e.g., Díaz-Campos 2004; Lord 2010; Nagle et al. 2016) to years (Piske et al. 2001). As noted by Simonet (2014), "The vast majority of work on interlingual or cross-linguistic phonetic influence in bilingualism does not explicitly distinguish between long-term and transient interference. Albeit implicitly, most studies explore features that are attributed to long-term interference" (p. 27). Yet more recent research has shown that the degree of cross-linguistic phonetic influence may also be subject to short-term variables.

2.2. Short-Term Sources of Cross-Linguistic Phonetic Influence

While the previous research detailed above highlights the notion that phonetic influence between a bilingual's two languages may occur as the result long-term factors, more recent research has begun to examine short-term sources of cross-linguistic phonetic influence. As Simonet (2014) notes, other authors use somewhat different terminology to refer to the same or similar phenomenon (p. 27). Grosjean (2012) differentiates between static and transient sources of cross-linguistic influence, while Paradis (1993) differentiates between competence and performance related cross-linguistic interference. For reasons noted in Simonet (2014), the long-term vs. short-term distinction is used henceforth. Here, *short-term* is used in contrast to long-term, broadly referring to situations or contexts in which phonetic production or perception is altered for a single speaker in response to immediate changes in the broader linguistic situation. These short-term factors may include a shift in the languages used in an interaction, also called the language context (e.g., Olson 2016), changes in external sociolinguistic factors, and even use of cognate tokens.

Several studies have begun to examine the potential of the language of a given interaction or experimental session as a source of short-term cross-linguistic phonetic influence. Simonet (2014), for example, examined the production of Catalan vowels by Spanish-Catalan bilinguals across two different sessions: unilingual and bilingual. In the unilingual session, all stimuli—utterances containing the target Catalan tokens—were drawn from Catalan, as were the session instructions. In contrast, stimuli in the bilingual session included utterances containing the target words from both Spanish and Catalan, in random order. It is important to note that the target tokens were embedded in meaningful utterances, and switches between Catalan and Spanish took place only at the utterance level. As such, the tokens under consideration were all non-switched tokens. Results from the production experiment revealed that Catalan vowels differed between the monolingual and bilingual sessions, with Catalan vowels becoming more Spanish-like in the bilingual session. Similarly, in a cued picture-naming paradigm, Olson (2013) compared the VOT of English and Spanish tokens produced in a monolingual session (i.e., 95% English—5% Spanish or 95% Spanish—5% English) with tokens produced in a bilingual session (i.e., 50% English—50% Spanish). Results showed differences in phonetic production of VOT depending on the nature of the session. Non-switched productions in the bilingual contexts, specifically in a participant's L1, shifted in the direction of the opposite language, with English VOTs becoming more Spanish-like and Spanish VOTs becoming more English-like. A similar effect may be seen in perception, in which the auditory context (i.e., English-like or Spanish-like acoustic features) surrounding a given ambiguous token serves to engage a language-specific perceptual system (Gonzales and Lotto 2013; for perceptual boundaries and language modes, see Casillas and Simonet (2018)). Further research has shown that while bilingual experimental sessions may foster cross-linguistic phonetic influence, the nature of this transfer may be dependent on a speaker's proficiency or dominance (Amengual 2018). Notably, Amengual (2018) found that bilinguals were more likely to experience cross-linguistic transfer resulting from a bilingual paradigm in their less dominant language.

These results from experimental paradigms also find some preliminary support from a more naturalistic paradigm. In her study of bilingual English-Arabic speaking children, Khattab (2002) collected naturalistic data in both English and Arabic-oriented language sessions. The results showed that while children clearly differentiated between the two phonetic systems, particularly with respect to /r/, English tokens produced during the Arabic sessions underwent a degree of phonetic transfer, becoming decidedly more Arabic-like. Subsequent analysis suggests that such transfer may relate to the language dominance of the interlocutor (Khattab 2009), with Arabic-accented English used during interactions with Arabic-dominant listeners. These results suggest that the use of two languages in the same interaction may serve to promote a degree of cross-linguistic influence.

Further evidence for cross-linguistic phonetic influence arising from the use of two languages in a single interaction can be seen in work on the phonetics of code-switching. *Code-switching* refers to the alternation between two or more languages or language varieties in a single discourse (e.g., Myers-Scotton 1997). As such, code-switching represents a clear point in an interaction in which both languages are simultaneously (or nearly simultaneously) activated and serves as a potential short-term source of cross-linguistic phonetic influence. Unlike the bilingual sessions detailed above (e.g., Khattab 2002; Olson 2013; Simonet 2014), which varied the ratio of the language used in a given session or block but examined non-switched productions, research focused on the phonetics of code-switching has generally focused on the potential for phonetic transfer at or near the point of switch. A growing body of research has begun to establish that code-switching impacts phonetic production at the segmental level, most notably inducing a degree of phonetic transfer (although, for a lack of transfer, see Grosjean and Miller (1994) and Muldner et al. (2019)). The exact nature of the phonetic influence found has varied. Several studies have found evidence of unidirectional transfer at the point of switch (Antoniou et al. 2011; Balukas and Koops 2015; Bullock et al. 2006), with Language A shifting in the direction of Language B, but Language B failing to show evidence of transfer. Other studies have found bidirectional transfer (Bullock and Toribio 2009; González-López 2012; Olson 2016; Schwartz et al. 2015), with Language A shifting in the direction of Language B and Language B shifting in the direction of Language A (for an account of unidirectional vs. bidirectional transfer, see Olson (2019)). This cross-linguistic phonetic influence has been found across a variety of paradigms, including naturalistic (e.g., Balukas and Koops 2015) and read speech (e.g., Antoniou et al. 2011), and for different types of code-switches (e.g., for single-word insertions, see Olson (2016); for alternational code-switching, see Bullock and Toribio (2009)). These shifts are largely phonetic in nature, rather than phonological, and bilinguals generally do not implement phonological categories of the opposite language. Thus, code-switching, which activates both of a bilingual's languages in a compressed timeframe, appears to serve as a short-term source for bilingual phonetic influence.

An additional case for short-term phonetic influence driven by linguistic factors can be seen in the production of cross-linguistic cognates. Cognates, words that have significant cross-linguistic overlap in meaning, phonology, and orthography (Amengual 2018), may result in the activation of both language systems, and as such represent a possible short-term source of cross-linguistic influence. Cognates have been shown to be produced with a degree of phonetic transfer. Amengual (2012), for example, showed that Spanish-English bilinguals produced longer (i.e., more English-like) VOTs in Spanish for cognate words than non-cognate words, a finding that held for heritage speakers (i.e., early bilinguals) and both late L1 English—L2 Spanish and L1 Spanish—L2 English bilinguals (for Spanish-Catalan bilinguals, see Amengual (2016)). Again, as cognates may activate both languages, they can be seen as a short-term source of cross-linguistic phonetic influence.

Changes in the language of the paradigm or interaction (e.g., Simonet 2014), code-switching (e.g., Olson 2016), and cognate status represent cases in which a short-term or immediate shift in the linguistic content of an interaction favors a degree of cross-linguistic phonetic influence. Yet, there is some evidence that non-linguistic changes in the external environment may also impact linguistic behavior. Hay and Drager (2010), for example, found that including region-specific objects in the experimental environment, such as a stuffed kangaroo (i.e., Australia-specific) or kiwi (New Zealand-specific), impacted vowel perception. The authors suggest that objects in the "ambient environment" can impact participant phonetic perception (p. 889). Other studies have found that visually salient characteristics of a speaker may influence phonetic perception (e.g., for intelligibility and accentedness, see Babel and Russell (2015)). For example, in a perceptual experiment, Koops et al. (2008) found that listeners' phonetic perception of stimuli reflecting an on-going, age-graded phonetic change (i.e., PIN-PEN unmerger in Houston, TX, USA) depended on perceived speaker age. Paralleling the change in progress in the local community, listeners were more likely to assume merged phonetic categories for older speakers and unmerged categories for younger speakers. The impact of such external factors on phonetic perception also extends to non-visible social information, such as supposed geographic origin (e.g., Niedzielski 1999). In these cases, it is not the linguistic content of an interaction or paradigm that shifts, but rather the surrounding environment and/or perceived interlocutor.

It is worth considering these short-term sources of cross-linguistic influence within the framework of a bilingual's language modes (e.g., Grosjean 1998, 2001, 2008). Bilinguals have the ability to operate along a linguistic continuum from operation entirely in Language A (i.e., monolingual mode) to operation in Language B (i.e., monolingual mode), including a variety of bilingual modes in which each of the two languages may be used to differing degrees. Language mode has been described in terms of the relative activation of each of the bilinguals two languages, with monolingual mode involving the activation of only (or predominantly) one language and a balanced bilingual mode involving the roughly equal activation of the two languages (Grosjean 2008). Grosjean (2008) notes that a speaker's language mode may be impacted by a variety of factors, including the "form and content of the message," the language act, the interlocutors, and the general situation of the interaction. All of these factors may be considered "short-term," in that they are subject to change from day-to-day or even interaction-to-interaction for a single bilingual speaker. Moreover, shifts in a speaker's (or listener's) position on the language mode continuum may impact their language production (or perception) patterns (e.g., Soares and Grosjean 1984). The previous findings of differing levels of cross-linguistic phonetic influence presented above can be reconceptualized as resulting from differing language modes. Short-term sources of cross-linguistic influence can be seen as variables that cause an immediate shift in a bilingual's position along the language mode continuum. Thus, short-term variables, including the language(s) of a given interaction or paradigm, the use of code-switching, and changes in the surrounding environment, may effectively serve to manipulate the relative activation (or suppression) of a bilingual's two languages, with more equal activation resulting in greater degrees of cross-linguistic phonetic influence.

2.3. Research Questions

Previous research has established that bilinguals, including late L2 learners, establish different phonetic norms for their two languages. Moreover, there appear to be both long-term (e.g., acquisition and immersion) and short-term (e.g., bilingual mode, code-switching) sources of phonetic cross-linguistic interaction. While long-term immersion in a given linguistic environment has been shown to impact the degree of cross-linguistic influence, the current study examines the potential for linguistic environment to act as a short-term source of cross-linguistic phonetic influence. That is, whether a change in linguistic environment results in an immediate shift in the degree of cross-linguistic phonetic influence. Two specific research questions are addressed:

> RQ1: Does a short-term change in the linguistic environment impact phonetic production? For this study, a short-term change in linguistic environment is operationalized as a single speaker moving from an English-dominant linguistic environment to a Spanish-dominant linguistic environment (or vice-versa).

Hypothesis 1. *Drawing on previous research that has shown an impact of both long-term and short-term sources of cross-linguistic phonetic influence, it was anticipated that a shift in linguistic environment would result in a corresponding shift in phonetic production. Specifically, L1 English—L2 Spanish speakers would produce Spanish tokens with more English-like phonetic features in an English-dominant linguistic environment and more Spanish-like phonetic features in a Spanish-dominant environment.*

> RQ2: Does proficiency in the L2 interact with linguistic environment?

Hypothesis 2. *Given the previous finding that as L2 experience and proficiency increase, L2 phonetic production shows less evidence of L1 to L2 cross-linguistic influence, it was anticipated that speakers with greater proficiency in the L2 may show smaller effects of a change in linguistic environment on their phonetic production.*

To assess the role of linguistic environment on phonetic production, the same participants were tested in both an English-dominant (Indiana, USA) and Spanish-dominant environment (Madrid, Spain). To focus squarely on the potential for linguistic environment to serve as a short-term source of cross-linguistic influence and limit the long-term effects of acquisition and immersion, participants were tested immediately prior to leaving one environment (i.e., less than 72 h pre-departure) and immediately upon arrival in the second environment (i.e., in the first 72 h in the new environment). The order of sessions was counterbalanced, such that one group was tested first in the English-dominant linguistic environment and then in the Spanish-dominant linguistic environment, and the other group received the opposite session order (i.e., Spanish-dominant environment then English-dominant linguistic environment).

Two different levels of phonetic analysis were conducted: an acoustic analysis of the voice onset time cue (VOT) and a global accent rating (GAR). These two measures were chosen to provide both a fine-grained measure of a relevant segment that differs cross-linguistically between English and Spanish (i.e., VOT), as well as a more global measure of perceived accent to capture potential shifts in other features (e.g., vowels, suprasegmental features, etc.) of production (i.e., GAR).

3. Methodology

3.1. Participants

Twenty English-speaking (L1) learners of Spanish (L2) participated in the current study. All participants were enrolled in a six-week immersive study abroad program in Madrid, Spain (Spanish-dominant environment) during the summers of either 2015 or 2018. The study abroad program included host family stays, classes at the local university, and several day trips to surrounding cultural sites, all of which were conducted in the L2. Participants were all learners at the intermediate to advanced level, enrolled in 3rd or 4th year university language courses. All participants gave informed consent prior to beginning the task and the protocol was approved by the Institutional Review Board at Purdue University (Protocol #: 1303013396).

Immediately prior to the main task, participants completed the Bilingual Language Profile (BLP; Birdsong et al. 2012). The BLP relies on self-assessments of a participant's language history, language use, language proficiency, and language attitudes in each of the relevant languages to provide a composite language score. All participants are native speakers of English, having learned English from birth (*M* age of acquisition = 0.0, *SD* = 0.0) and Spanish after the age of 5 (*M* = 12.1, *SD* = 3.0). Across all components, participants self-rated English as higher than Spanish (Table 1).

Table 1. Unweighted results of the Bilingual Language Profile (BLP) subcomponents.

Component	Scale [a]	English M (SD)	Spanish M (SD)
Language History	0–120	104.2 (8.7)	17.2 (8.4)
Language Use	0–50	46.6 (2.8)	3.3 (2.8)
Language Proficiency	0–24	23.8 (0.6)	14.2 (2.6)
Language Attitudes	0–24	23.4 (2.8)	15.0 (5.4)

[a] For each scale, higher ratings correspond to a more engagement with that component of a given language.

Following the BLP scoring procedures (for details, see Birdsong et al. (2012)), a composite language score was computed for each participant in each language, giving equal weight to each subcomponent. The possible composite language score, henceforth referred to as proficiency, ranges from 0, corresponding to no proficiency in the language, to 218, indicating high proficiency in the language. As expected, participants reported high language scores for English (*M* = 204.6, *SD* = 7.9). In contrast, participants reported lower and more varied Spanish language scores (*M* = 77.5, *SD* = 18.4). Figure 1 illustrates the distribution of participants with respect to their Spanish language score. All participants are considered to be English-dominant. All participants reported normal speech and hearing.

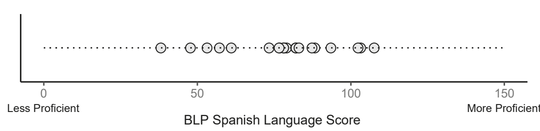


Figure 1. BLP Spanish language score by participant.

To assess the effect of linguistic environment, but counter-balance session order, participants (*N* = 20) were divided into two mutually exclusive groups. The first group (*n* = 12) was tested first in the English-dominant environment and then in the Spanish-dominant environment. The second group (*n* = 8) was tested first in the Spanish-dominant environment and then in the English-dominant environment. Again, to limit any possible long-term sources of cross-linguistic influence, participants were tested immediately prior to departure from Environment A (< 72 h) and immediately upon arrival in Environment B.[2]

[2] In short, one group of participants was tested en route to the host country of the study abroad program (i.e., prior to leaving Indiana, USA, and upon arrival in Madrid, Spain) and one group of participants was tested en route to the home country (i.e., prior to leaving Madrid, Spain, and upon arrival in Indiana, USA). While this presents a potential confound, such that one group may have had a six-week "advantage" by participating following six weeks of immersion in the host country, statistical analysis controlled for between-participant differences with the random effects structure, effectively comparing each participant to her or himself.

3.2. Stimuli

Stimuli for the read-aloud task were modified Spanish versions of utterances (*N* = 5) used in a global accent rating task (e.g., Flege 1988; Riney and Flege 1998). Relevant for the VOT analysis, embedded within these utterances were a number of tokens with word-initial voiceless stops. English and Spanish both employ a bi-partite distinction between voiceless and voiced stops in word initial position. VOT, defined as the temporal difference between the release of the oral closure and the onset of vocal fold vibration, has been shown to be a reliable cue to this voicing distinction (e.g., Lisker and Abramson 1964). Although both English and Spanish make use of this phonological distinction between voiceless and voiced phonemes, the phonetic cues differ. Specifically for voiceless stop consonants, English stops are produced with long-lag VOT (30–100 ms), while Spanish is produced with short-lag VOT (0–30 ms). Given this cross-linguistic difference, English-speaking learners of Spanish are tasked with acquiring and maintaining separate the Spanish-like short-lag VOT norms. A number of authors have noted that English-speaking learners of Spanish may produce Spanish voiceless stop consonants with English-like VOTs (e.g., Hammond 2001). Figure 2 shows the spectrogram and waveform for the word <*calle*> [kaʝe] 'street,' produced by a native English speaker (left) and native Spanish speaker (right). While the use of English-like VOT in Spanish voiceless stops is unlikely to cause issues of intelligibility (Lord 2005; Munro and Derwing 1995), it may impact the perception of speaker accentedness. Given both the gradient nature of VOT, and the previous evidence that bilinguals, including L2 learners, are able to effectively distinguish between English and Spanish VOT, VOT may serve as a sensitive measure of cross-linguistic influence between the two phonetic systems. A total of eight words contained word-initial voiceless stops, with the following distribution: /p/ = 2, /t/ = 2, /k/ = 4.

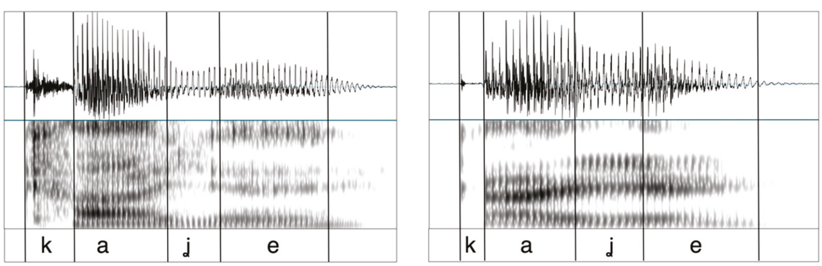


Figure 2. Spectrogram and waveform of <*calle*> 'street' produced by a native English speaker (**left**) and a native Spanish speaker (**right**). The difference in VOT is visible in the phoneme /k/.

Relevant for the GAR task, the five utterances contained a wide variety of segments (i.e., consonants and vowels). Several segments were included that differ significantly in English and Spanish phonetic production (e.g., word-initial <r>, which is produced as /ɹ/ in English and /r/ in Spanish), which could potentially serve as markers of English-accented Spanish. In addition, several segments were included that differ across different dialects of Spanish (e.g., <z>, <ci>, and <ce>, are produced as /s/ in most dialects of Spanish, but as /θ/ in Peninsular Spanish), which could potentially serve as markers of the local (Peninsular) dialect of Spanish.

Example (1) provides several sample stimuli. Target phonemes for the VOT analysis are underlined.

1. a. Te voy a leer este poema. 'I am going to read this poem to you.'
2. b. ¿Doblaste a la derecha en la calle principal? 'Did you turn to the left on the main street?'
3. c. Jaime comió los caramelos. 'Jaime ate the candies.'

3.3. Procedure

To assess the potential impact of linguistic environment, each speaker participated in two experimental sessions, each conducted in a different linguistic environment. One experimental session was conducted in an English-dominant linguistic environment (Indiana, USA), and one session was conducted in a Spanish-dominant linguistic environment (Madrid, Spain). In Indiana, USA, English is the home language of approximately 91.1% of the population, with Spanish spoken in the home by only 4.7% of the population (U.S. Census Bureau 2018). In Spain, 90% of the population speaks Spanish as a native language, while only 2.2% of the population speaks English as a native language (Instituto Nacional de Estadística 2019), although statistics were not available for the Community of Madrid. Moreover, both Indiana, USA, and Madrid, Spain, are considered to be "single language" environments in which little code-switching is present (Green and Abutalebi 2013). While the experience of each individual participant may vary, a finding captured by the language background questionnaire, the two environments are clearly distinguishable by the language of the broader environment.

Interaction with the experimenter was intentionally conducted in both languages in each session. The experimenter was a native speaker of midwestern American English who had spent several years living in Madrid, Spain and was proficient in the local Spanish dialect. Written instructions, provided before the start of the oral production task, were comprised of both English and Spanish. With the exception of the location of the two sessions, other experimental factors were maintained as equal as possible, using identical equipment, instructions, and consent forms and conducted with the same experimenter.

Following the collection of language background information, target utterances were presented visually using SuperLab v.5 (Cedrus Corporation 2015) and each utterance was repeated three times during each session. Utterances were recorded in a quiet room with a head-mounted microphone using Audacity v. 2.2.2 recording software. Both the instructions and the recording equipment were the same in each of the two different environments.

3.4. Voice Onset Time Analysis

A total of 960 tokens were considered in the initial VOT analysis (20 speakers × 8 tokens × 3 repetitions × 2 session = 960 tokens). Twenty-five tokens were classified as missing, and an additional 29 tokens were eliminated for a variety of speech errors (i.e., false start on target word) and recording errors (i.e., noisy recording). Lastly, outliers were eliminated (n = 5), defined as those tokens with VOT values greater than 3SD above and below the mean. A total of 901 tokens were included in the final VOT analysis.

VOT was defined as the temporal difference between the release of the oral closure and the onset of vibration of the vocal folds (e.g., Lisker and Abramson 1964). Tokens were measured using Praat (Boersma and Weenink 2018), with particular attention to the waveform. Tokens were coded blindly by a trained research assistant who was unaware of the linguistic environment in which the utterance was produced.

Statistical analysis was conducted using R statistical software (R Core Team 2013) and the lme4 package (Bates et al. 2015). Following recommendations by Barr et al. (2013), the maximal random effects structures that permitted model convergence were used. The significance criterion was set at $|t| > 2.00$. Power analysis was conducted with the simr package (Green and MacLeod 2016).

3.5. Global Accent Rating Analysis

Five native Spanish speakers from the target region (Madrid, Spain) were recruited as raters for the GAR task. All raters were native speakers of the target dialect, and using the BLP (Birdsong et al. 2012), were considered highly dominant in Spanish. The rating procedure was largely based on work by Riney and Flege (1998).

Two of the three repetitions per utterance were selected from each session for presentation to the native speakers. When possible, preference was given to the second and third repetitions of the stimuli. For productions containing speech errors, such as pauses or fillers, the first repetitions was substituted. Raters could listen to each presentation multiple times, if needed. The intensity of each utterance was scaled via script in Praat (Boersma and Weenink 2018) to 65 dB. The order of presentation was fully randomized. A total of 400 learner-produced utterances were selected for presentation to the native raters.

Raters were asked to provide accent ratings for each utterance on a 9-point Likert scale in which 1 corresponded to a "very strong non-native accent" and 9 a "native accent." Each accent rating was converted to a z-score on a by-rater basis to normalize for the different ranges of values used by each rater. A total of 2000 ratings were provided by native speakers (5 utterances × 2 repetitions × 2 sessions × 20 participants × 5 raters).

Statistical analysis was again conducted using R statistical software (R Core Team 2013) and the lme4 (Bates et al. 2015) and simr (Green and MacLeod 2016) packages. The maximal random effects structures that permitted model convergence were used (Barr et al. 2013).

4. Results

4.1. Voice Onset Time

An initial mixed effects model was conducted with VOT (ms) as the dependent variable and linguistic environment (English-dominant environment vs. Spanish-dominant environment) as the fixed effect. Random effects included participant and item (i.e., word), with random slopes and intercepts. Examination of a Q-Q plot confirmed that the residuals of the model were normally distributed. Contrary to the initial hypotheses, results from this initial model demonstrated no significant effect of linguistic environment on VOT, with similar VOTs produced in the English- (M = 44.2 ms, SD = 19.4 ms) and Spanish-dominant (M = 43.6 ms, SD = 18.8 ms) environments. The results for the fixed effects are available in Table 2 (for random effects and model equation, see Appendix A). Figure 3 shows the VOTs produced in each linguistic environment, separated by initial phoneme. Again, while we expect some differences in VOT across place of articulation (Cho and Ladefoged 1999), the key comparison is between VOT in the English and Spanish environments.

Table 2. Voice Onset Time (VOT) Model Fixed Effects.

	Estimate	Std. Error	*t*-Value	Lower 95%	Upper 95%	*d*	95% CI
Intercept (English Environment)	44.45	3.37	13.175	37.70	51.19		
Spanish Environment	0.22	1.85	0.119	−3.47	3.91	0.03	[−0.15, 0.22]

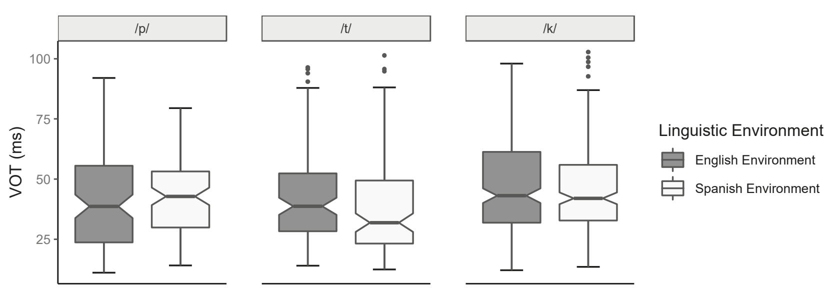


Figure 3. VOT (ms) by linguistic environment and place of articulation.

To ensure that the lack of a significant effect of linguistic environment on VOT production was not the result of an underpowered study, a power analysis was conducted using the simr package (Green and MacLeod 2016). Results of a simulation-based power analysis, with a medium effect size ($d = 0.5$) and based on 500 simulations, showed that the current study design surpassed the 80% power threshold (power for predictor linguistic environment = 99.8%, CI = [98.9, 100]). The outcome of the power analysis suggests that the lack of a significant effect of linguistic environment is not likely to be the result of an underpowered study design.

Related to the second research question, namely whether the effect of linguistic environment is conditioned by a given participant's proficiency in the target language, a second mixed effects model was conducted with the dependent variable of VOT. Fixed effects included linguistic environment (English-dominant environment vs. Spanish-dominant environment), proficiency, and their interaction. Proficiency was included as a continuous variable, with proficiency values determined by each participant's overall BLP language score for Spanish (Birdsong et al. 2012).[3] Random effects included participant and item, with random intercepts and random slopes by linguistic environment. More complex random effects structures, specifically random slopes by each of the two fixed effects, did not permit model convergence. Examination of a Q-Q plot confirmed that the residuals of the model were normally distributed. Results of this second model showed (see Table 3) that there was no significant effect of either linguistic environment or proficiency on VOT production. Moreover, there was no significant interaction between these two fixed effects, suggesting that linguistic environment was not a factor, regardless of a given participant's level of proficiency (for random effects and model equation, see Appendix B). Figure 4 illustrates these results. Again, lower VOT corresponds to more native-like pronunciation. Note that, while proficiency was included as a linear predictor in the model, Figure 4 grouped participants by relative proficiency for the purposes of visualization. Worth noting, the general expected trend is visible in Figure 4, with participants with higher proficiency in Spanish showing lower, more Spanish-like VOTs. However, it is clear that there is no effect of linguistic environment.[4]

Table 3. Voice Onset Time Model with Proficiency Fixed Effects.

	Estimate	Std. Error	*t*-Value	Lower 95%	Upper 95%
Intercept (English Environment)	58.53	13.36	4.380	31.80	85.26
Spanish Environment	1.41	7.37	0.191	−13.33	16.14
Proficiency	−0.18	0.17	−1.088	−0.52	0.15
Spanish Environment: Proficiency	−0.02	−0.02	−0.166	0.02	−0.05

3 Following suggestions by an anonymous reviewer, subsequent analysis was conducted with proficiency as a three-way categorical variable. Two different group cut-offs were considered. First, parallel to the subgroups in Figure 4, three approximately equal-sized proficiency groups were considered: low (n = 6), mid (n = 7), and high proficiency (n = 7). Second, three proficiency groups were identified using visual analysis of participant BLP Spanish score distributions: low (n = 5), mid (n = 12), and high proficiency (n = 3). Model comparison, following the procedure outlined below, showed that neither categorical approach to proficiency significantly improved model fit for either the VOT (equal sized groups: $\chi^2(4) = 1.380$, $p = 0.848$; unequal sized groups: $\chi^2(4) = 7.427$, $p = 0.115$) or the GAR analysis (equal sized groups: $\chi^2(4) = 3.517$, $p = 0.475$; unequal sized groups: $\chi^2(4) = 5.690$, $p = 0.224$) relative to a model without proficiency. As such, proficiency, regardless of operationalization, does not appear to significantly influence VOT or interaction with linguistic environment for this group of learners.

4 As the main goal of this project was to examine the effect of linguistic environment, the main analysis compares the productions of participants in two different linguistic environments. The two groups (i.e., US-Spain and Spain-US) served to counter-balance session order. As such, some group differences are possible as the Spain-US group was tested following six weeks in the Spanish linguistic environment. Addressing the possible effect of group, a subsequent model was conducted with linguistic environment and group as fixed effects,. Model comparison showed that the inclusion of group significantly improved model fit ($\chi^2(2) = 7.395$, $p = 0.025$). This is not unexpected, given that overall, the Spain-US group (M = 39.6 ms, SD = 18.3 ms) produced significantly shorter VOTs than the US-Spain group (M = 46.8 ms, SD = 19.1 ms), $t(787) = 5.693$, $p < 0.001$. To confirm that the impact of linguistic environment was similar for each group, separate models were conducted for each group. The model structure was parallel to the main model above. Results suggested that linguistic environment did not significantly impact VOT for either the US-Spain group ($b = -2.616$, $t = -1.136$) or the Spain-US group ($b = 4.799$, $t = 1.923$).

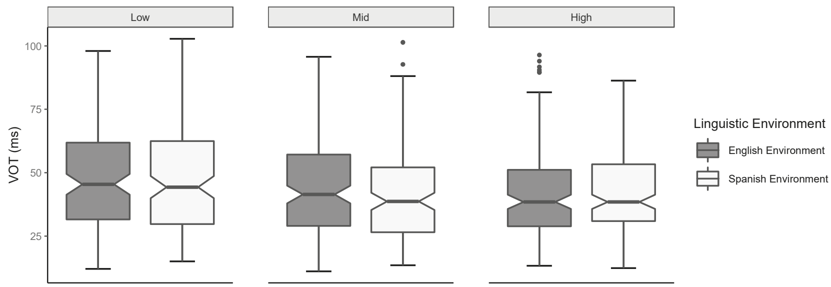


Figure 4. VOT (ms) by linguistic environment and proficiency. For the purposes of illustration only, participants were grouped into low (n = 6), mid (n = 7), and high proficiency (n = 7) groups by the continuous Spanish language component score of the BLP.

Finally, for completeness, a model comparison was done to assess the contribution of each of the variables of interest. Model comparison was conducted by comparing the model involving the two fixed effects (i.e., linguistic environment and proficiency), with the random effects structure previously detailed above, to submodels created by dropping one of the fixed effects but maintaining a similar random effects structure. Results of the model comparison show that there was no significant difference between the most complex model (log likelihood = −3706.4) and the submodel without the fixed effect of linguistic environment (log likelihood = −3706.4, $\chi^2(2)$ = 0.041, p = 0.979). Similarly, there was no difference between the most complex model and the submodel without the fixed effect of proficiency (log likelihood = −3707.2, $\chi^2(2)$ = 1.700, p = 0.428). As such, neither fixed effect contributed significantly to improving the model fit.

Taken as a whole, the results of the VOT analysis suggest that linguistic environment does not significantly impact the production of VOT. Moreover, for this group of learners, proficiency does not seem to be a relevant factor in the production of VOT.

4.2. Global Accent Rating

While the VOT analysis focuses on a specific segment, limited to only the voiceless stop consonants, the GAR analysis provides a more holistic metric of participant phonetic production. That is, while linguistic environment may not play a role in the production specifically of VOT, it is possible that other phonetic components, relevant to and noticeable by native speakers, may be modulated by environment.

Again, an initial mixed effects model was conducted with z-scored accent ratings as the dependent variable and linguistic environment as the fixed effect. Random effects included participant and item (i.e., utterance), with random slopes and intercepts. More complex random effects structures, particularly including rater as a random effect, did not permit model convergence. Again, each rating from the 9-point Likert scale was converted into a z-score on a by-rater basis. A visual analysis of the Q-Q plot confirmed that the residuals of the model were normally distributed. Results from this initial model closely parallel the results from the VOT analysis. Specifically, there was no significant impact of linguistic environment on accent ratings, with accent ratings for utterances produced in the English-dominant linguistic environment (M = -0.027, SD = 1.004) similar to those produced in the Spanish-dominant linguistic environment (M = 0.038, SD = 1.001). Full fixed effects results are seen in Table 4 (for random effects and model equation, see Appendix C).[5] Figure 5 illustrates the

[5] Parallel to the by-group analysis for VOT, mixed effect model was conducted on GAR with linguistic environment and group as fixed effects. Model comparison showed that the inclusion of group did not significantly improve model fit ($\chi^2(2)$ = 5.143,

global accent ratings by linguistic environment. Again, a higher accent rating corresponds to a more native-like production.

Table 4. Global Accent Rating Model Fixed Effects.

	Estimate	Std. Error	*t*-Value	Lower 95%	Upper 95%	*d*	95% CI
Intercept (English Environment)	−0.027	0.147	−0.185	−0.321	0.267		
Spanish Environment	0.065	0.082	0.795	−0.099	0.229	0.06	[−0.06, 0.19]

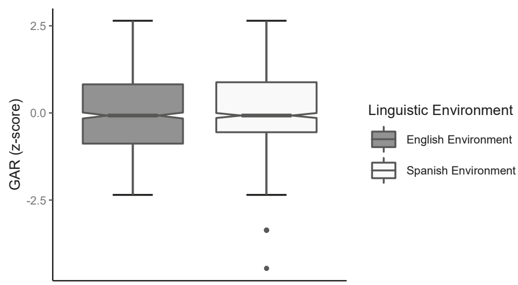


Figure 5. Global accent rating (z-scored) by linguistic environment.

To confirm that the lack of effect of linguistic environment on GAR was not due to an underpowered study, a power analysis was conducted using a simulation based-approach with the simr package (Green and MacLeod 2016). Results, based on 500 simulations with a medium effect size (d = 0.5), showed that the experiment exceeded the 80% threshold (power for predictor linguistic environment = 99.2%, CI = [97.9, 99.8]).

Considering the role of proficiency, a second model was conducted with linguistic environment and proficiency, as well as their interactions, as fixed effects. Participant and utterance were included as random effects, with random intercepts and slopes by linguistic environment, which was the maximal effects structure that permitted model convergence. Examination of a Q-Q plot confirmed that the residuals of the model were normally distributed. Results from the model with proficiency (for fixed effects, see Table 5; for random effects and model equation, see Appendix D) showed no significant effect of either linguistic environment or proficiency, and no significant interaction between the two fixed effects. Worth noting, the effect of proficiency trended in the expected direction, with participants who had higher self-rated Spanish language skills being rated as having more native-like accents.

Table 5. Global Accent Rating Model with Proficiency Fixed Effects.

	Estimate	Std. Error	*t*-Value	Lower 95%	Upper 95%
Intercept (English Environment)	−1.013	0.608	−1.666	−2.229	0.203
Spanish Environment	−0.006	0.287	−0.022	−0.580	0.568
Proficiency	0.012	0.008	1.667	−0.003	0.027
Spanish Environment: Proficiency	0.001	0.004	0.259	−0.006	0.008

Finally, a model comparison was conducted to examine the contribution of each of these fixed effects. The most complex model, with linguistic environment and proficiency as fixed effects, was compared to submodels created by dropping one of the fixed effects but maintaining the same random effects structure. Results from the model comparison showed that there was no significant

p = 0.076). As with VOT, results demonstrated that linguistic environment did not significantly impact the GAR for either the US-Spain group (b = 0.13, t = 1.169) or the Spain-US group (b = −0.03, t = −0.572).

difference between the complex model (log likelihood = −2407.7) and either the submodel without linguistic environment (log likelihood = −2408.0, $\chi^2(2) = 0.7496$, $p = 0.687$) or the submodel without proficiency (log likelihood = −2409.3.0, $\chi^2(2) = 3.315$, $p = 0.191$).

As with the analysis for VOT, the analysis from the GAR data suggests that linguistic environment did not impact native speaker ratings of learner productions. Moreover, while proficiency trended in the expected direction, it did not significantly interact with linguistic environment.

5. Discussion

The findings of this study add to the discussion on short-term sources of cross-linguistic influence and interaction. With respect to the first research question, and contrary to the initial hypothesis, the results showed that there was no significant effect of linguistic environment on cross-linguistic phonetic influence. The language of the broader community where an interaction took place had no relation to a bilingual's phonetic production. This result was found at both the fine-grained phonetic level, through an analysis of the VOT associated with Spanish voiceless stop consonants, as well as the more global level, as shown by the native speaker global accent ratings. Both VOT and perceived accent did not differ based on the linguistic environment of the session. Power analyses suggested that this lack of significant results is unlikely to be attributed to an underpowered study.

With respect to the second research question, namely the potential role of proficiency in modulating cross-linguistic phonetic influence, the results showed no significant role of proficiency for this particular group of speakers, and importantly, no interaction with the variable of linguistic environment. For all participants, regardless of proficiency, linguistic environment did not play a significant role in determining the degree of cross-linguistic influence or transfer. Again, this finding is contrary to the initial hypotheses.

The first hypothesis, specifically that linguistic environment would impact phonetic production, was driven by a robust body of research that has shown that bilingual phonetic production, and the degree of cross-linguistic phonetic interaction, is impacted by both long-term and short-term factors. Directly related to this hypothesis, we have seen that long-term immersion, either through immigration (Piske et al. 2001), travel (Sancier and Fowler 1997), or study abroad (Casillas 2020; Lord 2010; Solon and Long 2018), impacts phonetic production. Broadly, this research has shown that, over time, both a speaker's L2 and L1 (Bergmann et al. 2016; Major 1992; Stoehr et al. 2017) shift in the direction of the language of the broader community. Considering short-term sources of cross-linguistic phonetic influence, previous work has highlighted several short-term sources, including the language of a given interaction (e.g., Amengual 2018; Olson 2013; Simonet 2014), the use of code-switching (Antoniou et al. 2011; Balukas and Koops 2015; Bullock et al. 2006; Olson 2016), the presence of salient region-specific extra-linguistic cues in the interactional environment (Hay and Drager 2010), and visible (e.g., Babel and Russell 2015; Koops et al. 2008) and non-visible (Niedzielski 1999) social information about an interlocutor. Considering this line of research, it was anticipated that linguistic environment would impact production in the short-term, with phonetic targets shifting in the direction of the broader linguistic environment. Namely, it was anticipated that tokens produced in Madrid, Spain, would become more Spanish-like and tokens produced in Indiana, USA, would become more English-like. Given the lack of support for this hypothesis, it is worth considering several possible explanations.

One possible explanation for the lack of a short-term impact of linguistic environment is that the immediate local context of an interaction may be more relevant than the broader environment in which an interaction takes place. In much of the previous research on short-term sources of cross-linguistic phonetic influence, the source of the influence is present either in the interaction itself, either real (i.e., the language(s) required by the paradigm (e.g., Simonet 2014)) or imagined (i.e., visible or non-visible sociolinguistic cues (e.g., Babel and Russell 2015)), or is present in the physical environment that immediately surrounds participants (i.e., region-specific cues in the experimental setting (Hay and Drager 2010)). In each of these cases, the source of the short-term phonetic influence is in the speaker's immediate context. As such, the findings of the current study suggest a possible primacy of this immediate context relative to the broader linguistic environment. In short, the immediate context of the interaction is more relevant for phonetic production and perception than the broader linguistic environment, and short-term sources of cross-linguistic phonetic influence are local, rather than global. In the current study, the immediate context was maintained as similar as possible across the two experimental sessions. The same experimenter greeted participants, and the language of both the interaction with the experimenter (i.e., bilingual/code-switched interaction) and the written instructions were the same in both sessions. If there exists a primacy of the immediate context for short-term sources of cross-linguistic phonetic influence, these local characteristics of the interaction may have been more relevant than the linguistic environment of the broader community.

A second possible explanation for the lack of an impact of linguistic environment on phonetic production in the current study is related to the study population. While previous research has shown that L2-learners' phonetic productions are impacted by both long-term and short-term sources of cross-linguistic phonetic influence, the participants in the current study have relatively low proficiency in the L2. Directly related to their phonetic systems, the mean VOT values produced by these participants (for all productions M = 43.9 ms, SD = 19.1 ms) remain well outside the norms for native speakers (e.g., Lisker and Abramson 1964) and both early and late bilinguals (Amengual 2012). As such, it is possible that these participants are not sufficiently proficient in the L2 to respond to short-term sources of cross-linguistic influence at the phonetic level. The possible role of proficiency as a mitigating factor is echoed in previous work on L2 development during long-term engagement with a given language (e.g., study abroad), in which lower-proficiency speakers evidence less change in the L2 than higher-proficiency speakers during the immersion experience (for discussion of a threshold hypothesis in study abroad, see Lafford and Collentine (2006)). As such, a speaker's proficiency level may serve to modulate the impacts of short-term sources and effectively limit the role of linguistic environment in the current population. Moreover, the population for this study was fairly homogenous at the phonetic level, as illustrated by the minimal differences in VOT between the highest and lowest proficiency groups (see Figure 4; mean difference = 6.1 ms). This degree of homogeneity may further explain the lack of an impact of proficiency and the failure to support the second hypothesis.

Finally, it is worth returning to the language mode framework to provide an account for the role of both long-term and short-term sources of cross-linguistic influence. Grosjean (2001) provides a variety of factors that influence language mode: the interlocutors, the situation and physical location, the function of the language act, the type of stimuli and task, etc. It is worth noting that all of these factors can be considered as short-term sources of cross-linguistic influence and are subject to change within and between different interactions. Grosjean (2001) notes that language mode "concerns the level of activation of two languages" (p. 42), and the short-term variables may be the primary drivers of shifts in language mode. In contrast, the long-term sources of linguistic interaction, including acquisition, changes in proficiency, and immersion, are not listed as factors that impact language mode. As language mode has been described as a continuum, from monolingual operation in Language A to monolingual operation in Language B, short-term factors may serve to adjust a participant's position along their existing continuum. Long-term factors, such as proficiency, may serve to manipulate the nature of the endpoints of this continuum. Additional support for this interpretation comes from work in bilingual lexical access, notably from picture-naming tasks. This line of work has shown that short-term factors, such as the ratio of one language to another, impacts lexical access in both production (e.g., Gollan and Ferreira 2009; Olson 2015) and perception (e.g., Olson 2017), but that these effects are modulated by more long-term oriented factors like proficiency and language dominance (e.g., Olson 2015; Schwieter and Sunderman 2008). In terms of activation, long-term factors may ultimately manipulate a given language's baseline activation level or range of possible activation, while short-term factors manipulate the comparative level of activations of the two languages within their possible ranges.

Future Directions

Future research should continue to systematically examine the differential impacts of both long- and short-term sources of cross-linguistic phonetic influence. Building directly on the current study, and particularly in light of the failure of the results to support the initial hypotheses, future research may seek to expand upon the current study with a more heterogeneous population. Of particular interest may be to examine participants across a wide range of proficiencies, from learners to highly proficient early bilinguals, as they move through different linguistic environments. Moreover, developing a better understanding of other individual factors, such as participant's engagement with the local context (i.e., whether they maintain significant use of the L1 or immediately begin to engage in the L2), may serve to further our understanding of variability in cross-linguistic phonetic interaction. Second, it is acknowledged that the data set for the acoustic analysis is limited, both in terms of the number of tokens and the variety of features examined. Future research may seek to replicate these findings with a larger data set and a variety of phonetic features. For example, a more robust analysis across different places of articulation, precluded here by the size of the data set, may also be of interest. Notably, there appears to be some slight advantage for /t/ relative to the other places of articulation (see Figure 3), which would suggest that different phonemes may undergo different levels of cross-linguistic influence (for a discussion of potential differences in the perceptual prominence of voiceless stops by place of articulation, see Ruch and Peters (2016, p. 28)). Furthermore, research may seek to disentangle the possible effects of the immediate interactional context and the broader environment of that context, exploring the possible notion of a primacy of immediate or local factors as short-term sources of cross-linguistic influence.

6. Conclusions

This study examined the potential for linguistic environment, conceptualized as the language norms of the broader community in which an interaction or experiment takes place, to serve as a short-term source of cross-linguistic influence. To assess the role of linguistic environment, bilinguals (i.e., English speaking learners of Spanish) produced Spanish utterances in two sessions: an English-dominant linguistic environment (Indiana, USA) and a Spanish-dominant linguistic environment (Madrid, Spain). Productions were analyzed at the fine-grained acoustic level, though an acoustic analysis of voice onset time, as well as more holistically through native speaker global accent ratings. Results showed that linguistic environment did not significantly impact either measure of phonetic production. Moreover, there was no interaction of proficiency with linguistic environment, suggesting that the linguistic environment was not a relevant factor, regardless of participant proficiency in their second language.

The current findings, notably the lack of an impact of the broader linguistic environment in determining cross-linguistic phonetic influence, may suggest a primacy of the local factors (i.e., characteristics of the interaction and the immediately surrounding area) over broader, global factors (i.e., the linguistic environment of the broader community surrounding the interaction) as sources of short-term cross-linguistic interaction. Further research is needed to confirm these results and continue to explore the role of both long-term and short-term sources of cross-linguistic phonetic influence.

Funding: This research was funded in part by the Purdue Research Foundation.

Acknowledgments: I am grateful for the technical work of Serae Neidigh on this project. All errors are my own.

Conflicts of Interest: The author declares no conflict of interest. The funding sponsors had no role in the design of the study; in the data collection, analyses, or interpretation of the data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A

Table A1. Voice Onset Time Model Random Effects.

Participant	Variance	Std. Dev.	Corr.
Intercept	172.98	13.15	
Spanish Environment	34.33	5.59	−0.36
Item	**Variance**	**Std. Dev.**	**Corr.**
Intercept	15.95	3.99	
Spanish Environment	5.36	2.31	0.09

Equation:

$$VOT_{ijk} = \beta_0 + \beta_1 * I\Big(Environment_{ijk} = B\Big) + Participant_j + Item_k + \epsilon_{ijk}$$

Appendix B

Table A2. Voice Onset Time Model with Proficiency Random Effects.

Participant	Variance	Std. Dev.	Corr.
Intercept	171.17	13.08	
Spanish Environment	37.17	6.10	−0.39
Item	**Variance**	**Std. Dev.**	**Corr.**
Intercept	15.95	3.99	
Spanish Environment	5.37	2.32	0.09

Equation:

$$\begin{aligned} VOT_{ijk} = {} & \beta_0 + \beta_1 * I\Big(Environment_{ijk} = B\Big) + \beta_2 * Proficiency + \beta_3 \\ & * I\Big(Environment_{ijk} = B\Big) * Proficiency + \gamma_j * I(Environment_{ijk} \\ & = B) + \eta_k * I\Big(Environment_{ijk} = B\Big) + Participant_j + Item_k + \epsilon_{ijk} \end{aligned}$$

Appendix C

Table A3. Global Accent Rating Model Random Effects.

Participant	Variance	Std. Dev.	Corr.
Intercept	0.399	0.632	
Spanish Environment	0.053	0.230	−0.17
Item	**Variance**	**Std. Dev.**	**Corr.**
Intercept	0.005	0.072	
Spanish Environment	0.014	0.119	−0.2

Equation:

$$GAR_{ijk} = \beta_0 + \beta_1 * I\Big(Environment_{ijk} = B\Big) + Participant_j + Item_k + \epsilon_{ijk}$$

Appendix D

Table A4. Global Accent Rating Model with Proficiency Random Effects.

Participant	Variance	Std. Dev.	Corr.
Intercept	0.364	0.604	
Spanish Environment	0.056	0.238	−0.22
Item	**Variance**	**Std. Dev.**	**Corr.**
Intercept	0.005	0.072	
Spanish Environment	0.014	0.119	−0.2

Equation:

$$\begin{aligned} GAR_{ijk} &= \beta_0 + \beta_1 * I\big(Environment_{ijk} = B\big) + \beta_2 * Proficiency + \beta_3 \\ &* I\big(Environment_{ijk} = B\big) * Proficiency + \gamma_j * I(Environment_{ijk} \\ &= B) + \eta_k * I\big(Environment_{ijk} = B\big) + Participant_j + Item_k + \epsilon_{ijk} \end{aligned}$$

References

Amengual, Mark. 2012. Interlingual influence in bilingual speech: Cognate status effect in a continuum of bilingualism. *Bilingualism: Language and Cognition* 15: 517–30. [CrossRef]

Amengual, Mark. 2016. Cross-linguistic influence in the bilingual mental lexicon: Evidence of cognate effects in the phonetic production and processing of a vowel contrast. *Frontiers in Psychology* 7: 617. [CrossRef]

Amengual, Mark. 2018. Asymmetrical interlingual influence in the production of Spanish and English laterals as a result of competing activation in bilingual language processing. *Journal of Phonetics* 69: 12–28. [CrossRef]

Antoniou, Mark, Catherine T. Best, Michael D. Tyler, and Christian Kroos. 2011. Inter-language interference in VOT production by L2-dominant bilinguals: Asymmetries in phonetic code-switching. *Journal of Phonetics* 39: 558–70. [CrossRef] [PubMed]

Babel, Molly, and Jamie Russell. 2015. Expectations and speech intelligibility. *The Journal of the Acoustical Society of America* 137: 2823–33. [CrossRef] [PubMed]

Balukas, Colleen, and Christian Koops. 2015. Spanish-English bilingual voice onset time in spontaneous code-switching. *International Journal of Bilingualism* 19: 423–43. [CrossRef]

Barr, Dale, Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68: 255–78. [CrossRef]

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67: 1–48. [CrossRef]

Bergmann, Christopher, Amber Nota, Simone A. Sprenger, and Monika S. Schmid. 2016. L2 immersion causes non-native-like L1 pronunciation in German attriters. *Journal of Phonetics* 58: 71–86. [CrossRef]

Best, Catherine T., and Michael Tyler. 2007. Nonnative and second-language speech perception. In *Language Experience in Second Language Speech Learning: In Honour of James Emil Flege*. Edited by Murray J. Munro and Ocke-Schwen Bohn. Amsterdam: John Benjamins, pp. 13–34.

Birdsong, David, Libby M. Gertken, and Mark Amengual. 2012. *Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism*. Austin: COERLL, University of Texas at Austin. Available online: http://sites.la.utexas.edu/bilingual (accessed on 22 October 2020).

Boersma, Paul, and David Weenink. 2018. Praat: Doing Phonetics by Computer. Version 6.0.37. Available online: http://www.praat.org (accessed on 22 October 2020).

Bullock, Barbara E., and Almeida Jacqueline Toribio. 2009. Trying to hit a moving target: On the sociophonetics of code-switching. In *Multidisciplinary Approaches to Code Switching*. Edited by Ludmilla Isurin, Donald Winford and Kees de Bot. Amsterdam: John Benjamin, pp. 189–206.

Bullock, Barbara E., Almeida Jacqueline Toribio, Verónica González, and Amanda Dalola. 2006. Language dominance and performance outcomes in bilingual pronunciation. In *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference*. Edited by Mary Grantham O'Brien, Christine Shea and John Archibald. Somerville: Cascadilla Proceedings Project, pp. 9–16.

Caramazza, Alfonso, Grace H. Yeni-Komshian, Edgar B. Zurif, and Ettore Carbone. 1973. The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *The Journal of the Acoustical Society of America* 54: 421–28. [CrossRef]

Casillas, Joseph V. 2020. The longitudinal development of fine-phonetic detail: Stop production in a domestic immersion program. *Language Learning* 70: 768–806. [CrossRef]

Casillas, Joseph V., and Miquel Simonet. 2018. Perceptual categorization and bilingual language modes: Assessing the double phonemic boundary in early and late bilinguals. *Journal of Phonetics* 71: 51–64. [CrossRef]

Cedrus Corporation. 2015. *SuperLab Pro (Version 5.0.5) [Computer Program]*. San Pedro: Cedrus Corporation.

Cho, Taehong, and Peter Ladefoged. 1999. Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics* 27: 207–29. [CrossRef]

Díaz-Campos, Manuel. 2004. Context of learning in the acquisition of Spanish second language phonology. *Studies in Second Language Acquisition* 26: 249–74. [CrossRef]

Flege, James Emil. 1987. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15: 47–65. [CrossRef]

Flege, James Emil. 1988. Factors affecting degree of perceived foreign accent in English sentences. *The Journal of the Acoustical Society of America* 84: 70–79. [CrossRef]

Flege, James Emil. 1991. Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *The Journal of the Acoustical Society of America* 89: 395–411. [CrossRef]

Flege, James Emil. 1995. Second language speech learning: Theory, findings, and problems. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Edited by Winifred Strange. Timonium: York Press, pp. 233–77.

Flege, James Emil, and Wieke Eefting. 1987. Production and perception of English stops by native Spanish speakers. *Journal of Phonetics* 15: 67–83. [CrossRef]

Flege, James Emil, and Robert Port. 1981. Cross-language phonetic interference: Arabic to English. *Language and Speech* 24: 125–46. [CrossRef]

Flege, James Emil, Murray J. Munro, and Ian R. A. MacKay. 1995. Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America* 97: 3125–34. [CrossRef]

Flege, James Emil, Ian R. A. MacKay, and Diane Meador. 1999. Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America* 106: 2973–87. [CrossRef]

Fowler, Carol A., Valery Sramko, David J. Ostry, Sarah A. Rowland, and Pierre Hallé. 2008. Cross language phonetic influences on the speech of French–English bilinguals. *Journal of Phonetics* 36: 649–63. [CrossRef] [PubMed]

Gollan, Tamar H., and Victor S. Ferreira. 2009. Should I stay or should I switch? A cost–benefit analysis of voluntary language switching in young and aging bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 35: 640–65. [CrossRef]

Gonzales, Kalim, and Andrew Lotto. 2013. A *Barfi*, un *Pafri*: Bilinuals' pseudoword identifications support language-specific phonetic systems. *Psychological Science* 24: 2135–42. [CrossRef] [PubMed]

González-López, Verónica. 2012. Spanish and English word-initial voiceless stop production in code-switched vs. monolingual structures. *Second Language Research* 28: 243–63. [CrossRef]

Green, David W., and Jubin Abutalebi. 2013. Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology* 25: 515–30. [CrossRef] [PubMed]

Green, Peter, and Catriona J. MacLeod. 2016. SIMR: An R package for power analysis of generalised linear mixed models by simulation. *Methods in Ecology and Evolution* 7: 493–98. [CrossRef]

Grosjean, François. 1998. Studying bilinguals; methodological and conceptual issues. *Bilingualism: Language and Cognition* 1: 131–49. [CrossRef]

Grosjean, François. 2001. The bilingual's language modes. In *One Mind, Two Languages: Bilingual Language Processing*. Edited by Janet Nicol. Oxford: Oxford University Press, pp. 1–22.

Grosjean, François. 2008. *Studying Bilinguals*. Oxford: Oxford University Press.

Grosjean, François. 2012. An attempt to isolate, and then differentiate, transfer and interference. *International Journal of Bilingualism* 16: 11–21. [CrossRef]

Grosjean, François, and Joanne L. Miller. 1994. Going in and out of languages: An example of bilingual flexibility. *Psychological Science* 5: 201–6. [CrossRef]

Guion, Susan G., Tetsuo Harada, and J. J. Clark. 2004. Early and late Spanish–English bilinguals' acquisition of English word stress patterns. *Bilingualism: Language and Cognition* 7: 207–26. [CrossRef]

Hammond, Robert M. 2001. *The Sounds of Spanish: Analysis and Application (with Special Reference to American English)*. Somerville: Cascadilla Press.

Hay, Jennifer, and Katie Drager. 2010. Stuffed toys and speech perception. *Linguistics* 48: 865–892. [CrossRef]

Instituto Nacional de Estadística. 2019. Lenguas Maternas: Personas entre 18 y 64 años de edad según las lenguagas maternas más frequentes que tienen, por característica sociodemográfica. Available online: http://www.ine.es/jaxi/tabla.do?type=pcaxis&path=/t13/p459/a2016/p01/l0/&file=01016.px (accessed on 22 October 2020).

Jarvis, Scott, and Aneta Pavlenko. 2008. *Crosslinguistic Influence in Language and Cognition*. New York: Routledge.

Khattab, Ghada. 2002. /r/ production in English and Arabic bilingual and monolingual speakers. In *Leeds Working Papers in Linguistics and Phonetics 9*. Edited by Diane Nelson. Leeds: Leeds University, pp. 91–129.

Khattab, Ghada. 2009. Phonetc accommodation in children's code switching. In *The Cambridge Handbook of Linguistic Code Switching*. Edited by Barbara Bullock and Almeida Jacqueline Toribio. Cambridge: Cambridge University, pp. 142–60.

Koops, Christian, Elizabeth Gentry, and Andrew Pantos. 2008. The effect of perceived speaker age on the perception of PIN and PEN vowels in Houston, Texas. *University of Pennsylvania Working Papers in Linguistics* 14: 93–101.

Kuhl, Patricia K. 1992. Infants' perception and representation of speech: Development of a new theory. In *Proceedings of the International Conference on Spoken Language Processing*. Edited by John J. Ohala, Terrance M. Neary, Bruce L. Derwing, Megan H. Hodge and Grace E. Wiebe. Edmonton: University of Alberta, pp. 3–10.

Kuhl, Patricia K. 1993a. Infant speech perception: A window on psycholinguistic development. *International Journal of Psycholinguistics* 9: 33–56.

Kuhl, Patricia. K. 1993b. Innate predispositions and the effects of experience in Speech Perception. The Native Language Magnet Theory. In *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*. Edited by Bénédicte de Boysson-Bardies, Scania de Schonen, Peter Jusczyk, Peter McNeilage and John Morton. Dordrecht: Kluwer Academic Publishers, pp. 259–74.

Lafford, Barbara, and Joseph Collentine. 2006. The effects of study abroad and classroom contexts on the acquisition of Spanish as a second language. In *The Art Of Teaching Spanish: Second Language Acquisition: From Research To Praxis*. Edited by Barbara Lafford and Rafael Salaberry. Washington: Georgetown University Press, pp. 103–26.

Lee, Junkyu, Juhyun Jang, and Luke Plonsky. 2015. The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics* 36: 345–66. [CrossRef]

Lisker, Leigh, and Arthur S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20: 384–422. [CrossRef]

Lord, Gillian. 2005. (How) can we teach foreign language pronunciation? On the effects of a Spanish phonetics course. *Hispania* 88: 557–67. [CrossRef]

Lord, Gillian. 2010. The combined effects of immersion and instruction on second language pronunciation. *Foreign Language Annals* 43: 488–503. [CrossRef]

MacLeod, Andrea, and Carol Stoel-Gammon. 2005. Are bilinguals different? What VOT tells us about simultaneous bilinguals. *Journal of Multilingual Communication Disorders* 3: 118–27. [CrossRef]

Major, Roy C. 1992. Losing English as a first language. *The Modern Language Journal* 76: 190–208. [CrossRef]

Muldner, Kasia, Leah Hoiting, Leyna Sanger, Lev Blumenfeld, and Ida Toivonen. 2019. The phonetics of code-switched vowels. *International Journal of Bilingualism* 23: 37–52. [CrossRef]

Munro, Murray J., and Tracey M. Derwing. 1995. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* 45: 73–97. [CrossRef]

Myers-Scotton, Carol. 1997. *Duelling Languages: Grammatical Structure in Codeswitching*. Oxford: Oxford University Press.

Nagle, Charles L., Colleen Moorman, and Cristina Sanz. 2016. Disentangling research on study abroad and pronunciation: Methodological and programmatic considerations. In *Handbook of Research on Study Abroad Programs and Outbound Mobility*. Edited by Donna M. Velliaris and Deb Coleman-George. Hershey: IGI Global, pp. 673–95.

Niedzielski, Nancy. 1999. The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology* 18: 62–85. [CrossRef]

Olson, Daniel J. 2013. Bilingual language switching and selection at the phonetic level: Asymmetrical transfer in VOT production. *Journal of Phonetics* 41: 407–20. [CrossRef]

Olson, Daniel J. 2015. The gradient effect of context on language switching and lexical access in bilingual production. *Applied Psycholinguistics* 37: 725–56. [CrossRef]

Olson, Daniel J. 2016. The Role of code-switching and language context in bilingual phonetic transfer. *Journal of the International Phonetic Association* 46: 263–85. [CrossRef]

Olson, Daniel J. 2017. Bilingual language switching costs in auditory comprehension. *Language, Cognition and Neuroscience* 32: 494–513. [CrossRef]

Olson, Daniel J. 2019. Phonological processes across word and language boundaries: Evidence from code-switching. *Journal of Phonetics* 77: 1–16. [CrossRef]

Paradis, Michel. 1993. Linguistic, psycholinguistic, and neurolinguistic aspects of 'interference' in bilingual speakers: The Activation Threshold Hypothesis. *International Journal of Psycholinguistics* 9: 133–45.

Piske, Thorsten, Ian R. A. MacKay, and James Emil Flege. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29: 191–215. [CrossRef]

Piske, Thorsten, James Emil Flege, Ian R. A. MacKay, and Diane Meador. 2002. The production of English vowels by fluent early and late Italian-English bilinguals. *Phonetica* 59: 49–71. [CrossRef] [PubMed]

R Core Team. 2013. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, Available online: http://www.R-project.org/ (accessed on 22 October 2020).

Riney, Timothy J., and James E. Flege. 1998. Changes over time in global foreign accent and liquid identifiability and accuracy. *Studies in Second Language Acquisition* 20: 213–43. [CrossRef]

Ruch, Hanna, and Sandra Peters. 2016. On the origin of post-aspirated stops: Production and perception of/s/+ voiceless stop sequences in Andalusian Spanish. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 7: 1–36. [CrossRef]

Sancier, Michele L., and Carol A. Fowler. 1997. Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics* 25: 421–36. [CrossRef]

Schmid, Monika S., Steven Gilbers, and Amber Nota. 2014. Ultimate attainment in late second language acquisition: Phonetic and grammatical challenges in advanced Dutch–English bilingualism. *Second Language Research* 30: 129–57. [CrossRef]

Schwartz, Geoffrey, Anna Balas, and Arkadiusz Rojczyk. 2015. Phonological factors affecting L1 phonetic realization of proficient Polish users of English. *Research in Language* 13: 181–98. [CrossRef]

Schwieter, John W., and Gretchen Sunderman. 2008. Language switching in bilingual speech production: In search of the language-specific selection mechanism. *The Mental Lexicon* 3: 214–38. [CrossRef]

Simon, Ellen. 2009. Acquiring a new second language contrast: An analysis of the English laryngeal system of native speakers of Dutch. *Second Language Research* 25: 377–408. [CrossRef]

Simonet, Miquel. 2014. Phonetic consequences of dynamic cross-linguistic interference in proficient bilinguals. *Journal of Phonetics* 43: 26–37. [CrossRef]

Soares, Carlos, and François Grosjean. 1984. Bilinguals in a monolingual and a bilingual speech mode: The effect on lexical access. *Memory and Cognition* 12: 380–86. [CrossRef]

Solon, Megan, and Avizia Y. Long. 2018. Acquisition of phonetics and phonology abroad: What we know and how. In *The Routledge Handbook of Study Abroad: Research and Practice*. Edited by Cristina Sanz and Alfonso Morales-Front. New York: Routledge, pp. 71–85.

Stoehr, Antje, Titia Benders, Janet G. Van Hell, and Paula Fikkert. 2017. Second language attainment and first language attrition: The case of VOT in immersed Dutch–German late bilinguals. *Second Language Research* 33: 483–518. [CrossRef] [PubMed]

U.S. Census Bureau. 2018. ACS 1-Year Estimates Subject Tables: Languages Spoken at Home. Available online: https://data.census.gov/cedsci/table?q=Language%20Indiana&tid=ACSST1Y2018.S1601&hidePreview=false (accessed on 22 October 2020).

languages

MDPI

Article

Does Teaching Your Native Language Abroad Increase L1 Attrition of Speech? The Case of Spaniards in the United Kingdom

Robert Mayr [1,*], David Sánchez [1] and Ineke Mennen [2]

1 Centre for Speech and Language Therapy and Hearing Science, Cardiff Metropolitan University, Cardiff CF5 2YB, UK; davidjcox86@gmail.com
2 Department of English Studies, University of Graz, 8010 Graz, Austria; ineke.mennen@uni-graz.at
* Correspondence: rmayr@cardiffmet.ac.uk

Received: 30 August 2020; Accepted: 16 October 2020; Published: 22 October 2020

Abstract: The present study examines the perceived L1 accent of two groups of native Spaniards in the United Kingdom, Spanish teachers, and non-teachers, alongside monolingual controls in Spain. While the bilingual groups were carefully matched on a range of background variables, the teachers used Spanish significantly more at work where they constantly need to co-activate it alongside English. This allowed us to test the relative effect of reduced L1 use and dual language activation in first language attrition directly. To obtain global accentedness ratings, monolingual native Spanish listeners living in Spain participated in an online perception experiment in which they rated short speech samples extracted from a picture-based narrative produced by each speaker in terms of their perceived nativeness, and indicated which features they associated with non-nativeness. The results revealed significantly greater foreign-accent ratings for teachers than non-teachers and monolinguals, but no difference between the latter two. Non-native speech was associated with a range of segmental and suprasegmental features. These results suggest that language teachers who teach their L1 in an L2-speaking environment may be particularly prone to L1 attrition since they need to co-activate both their languages in professional settings and are regularly exposed to non-native speech from L2 learners.

Keywords: L1 attrition; speech; foreign accent; accent perception; Spanish; English; bilingual; teacher

1. Introduction

A growing body of research on speech development in early and late bilinguals has documented changes occurring in a speaker's first language (L1) pronunciation that affect areas, such as vowels, consonants, and intonation patterns (e.g., (de Leeuw 2019; Fowler et al. 2008; Mayr et al. 2019; Mennen 2004; Nodari et al. 2019)). Such changes can take place rapidly, affect novice second language (L2) learners (Chang 2012, 2013; Kartushina et al. 2016a), and may be fully (Kartushina and Martin 2019) or partially (Chang 2019) reversed. Alternatively, they may occur over time in proficient L2 learners who are long-term residents in an L2-speaking environment (e.g., (de Leeuw et al. 2018a; Dmitrieva 2019; Mayr et al. 2012; Ulbrich and Ordin 2014)). Only the latter scenario is usually referred to as "L1 attrition", that is, the non-pathological and non-age related decrease in an individual's proficiency in a previously learnt language ((Köpke and Schmid 2004; Schmid 2010), but see (Schmid and Köpke 2017) for a broader definition).

While L1 attrition of speech has been widely documented, not all individuals who are long-term residents in an L2-speaking environment exhibit observable changes to their native accent (de Leeuw et al. 2018a; Major 1992; Mennen 2004). The specific factors that facilitate or hinder attrition of speech are, however, still poorly understood, and few studies have systematically investigated relevant predictor variables (but see (Hopp and Schmid 2013)). One of the ongoing debates in this context is whether attrition is predominantly caused by reduced L1 use or by cross-linguistic interactions arising from contexts of dual language activation (de Leeuw et al. 2010; Schmid 2007; Stoehr et al. 2017). The present study aims to contribute to this issue by examining the perceived L1 accent of two groups of native Spanish speakers in the United Kingdom: (1) Spanish language teachers, who use their L1 regularly in professional settings and frequently need to switch between Spanish and English, and (2) non-teachers, who virtually never use the L1 in the workplace, with the two groups exhibiting similar use of the L1 in social situations. This design allowed us to test the role of low L1 use and regular dual language activation in L1 attrition of speech directly.

1.1. Plasticity in Native Speech, Phonetic Drift and L1 Attrition

Research into late bilingualism has until recently been primarily concerned with L2 acquisition where prevailing notions have been that a critical period (Lenneberg 1967) and processes of fossilization (Selinker 1972) constrain ultimate attainment in the L2. Whether this putative end state is maturationally constrained or conditioned by increasing entrenchment is still subject to ongoing debates (e.g., (Bylund et al. 2013; Piske et al. 2001); nonetheless, traditional perspectives on bilingualism have largely ignored the L1, assuming it to be stable and unlikely to undergo significant development (e.g., (Gregg 2010)).

Such suggestions, however, are not supported by empirical findings which show that bilinguals' L1 speech patterns typically differ from those of monolinguals (see (Kartushina et al. 2016b) for an overview, and below). Moreover, they are at odds with a holistic view of bilingualism, which argues that the L1 and L2 do not exist in isolation but constantly interact with each other (Grosjean 1989). In line with this account, the Speech Learning Model (SLM) (Flege 1995; Flege and Bohn 2020) posits that the L1 and L2 share a common phonological space and influence each other, which may lead to cross-linguistic assimilation and dissimilation patterns, both of which will differ from those of monolinguals. Moreover, the SLM claims that the more experienced an L2 learner is, the greater the effect of the L2 will be on the L1 (Flege 1995).

A static view of the L1 has also been challenged by the advocates of Dynamic Systems Theory (e.g., (de Bot et al. 2007)). According to this account, language constitutes a system with multiple components that are continually in a state of flux. These components are interconnected and sensitive to feedback, both from internal stimuli (i.e., other components within the system), and social and environmental factors. Thus, native speech patterns are dynamic and subject to change throughout the lifespan. Indeed, there is widespread evidence from longitudinal studies that show that even monolinguals modify their L1 accent in response to changes in the norms of their speech community (Harrington 2006; Harrington et al. 2000; Sankoff and Blondeau 2007). Amongst these, a particularly well-known example is the work of Harrington and his associates which showed systematic changes over several decades in the Queen's vowel realizations during her annual Christmas address.

Changes in L1 accent have also been widely documented in longitudinal work on bilinguals. For example, Sancier and Fowler (1997) present the case study of a Brazilian Portuguese-English bilingual who regularly travelled between Brazil and the United States (see also (Tobin et al. 2017) for a recent extension to Spanish-English bilinguals). They found that her voice onset time (VOT) values in both languages were longer after several months in the United States and shorter after months in Brazil, a change to which native Portuguese listeners were receptive. The authors ascribe the observed variation to what they call a "gestural drift" (more recently "phonetic drift" (Chang 2012, 2013), suggesting that L1 phones begin to adopt characteristics of the ambient language as a result of their similarity to L2 phones, the speakers' propensity to unintentionally imitate what they hear, and the effect of recency on memory. Since phonetic drifts of this kind do not coincide with a decline in L1 proficiency, they are not considered instances of attrition (see also (Chang 2012, 2013; Kartushina et al. 2016a)).

In contrast, an extensive body of literature has documented pervasive changes in L1 accent in bilinguals who are long-term residents in an L2-speaking environment. At the phonetic level, such instances of L1 attrition have been shown to affect the production of VOT in plosives (Flege 1987; Major 1992; Mayr et al. 2012; Stoehr et al. 2017), formant frequencies in vowels (Bergmann et al. 2016; Guion 2003; Mayr et al. 2012), laterals (de Leeuw et al. 2013; de Leeuw 2019) and rhotics (de Leeuw et al. 2018b; Ulbrich and Ordin 2014), and the realization of tonal alignment (de Leeuw et al. 2012; Mennen 2004). Attrition has also been shown to affect L1 perception (Ahn et al. 2017; Dmitrieva 2019) and may result in the neutralization of native phonological contrasts (Cho and Lee 2016; de Leeuw et al. 2018a).

Moreover, there is ample evidence that listeners are receptive to changes in L1 accent and may perceive speakers as foreign accented in their native language (Bergmann et al. 2016; de Leeuw et al. 2010; Hopp and Schmid 2013). For example, de de Leeuw et al. (2010) examined the global foreign accent in the L1 of native German speakers who were long-term residents in Anglophone Canada or the Netherlands. The results revealed that they were perceived to be significantly less native-like than native control speakers in Germany, irrespective of geographical setting. Similarly, the native German speakers in Anglophone North America in Bergmann et al. (2016) were perceived as significantly less native-like in their L1 than control speakers in Germany, with 40% of attriters rated below the monolingual range.

Together, the extant literature hence suggests that L1 attrition of speech is widespread and may be observed both in the productions of bilinguals and in their global foreign accent ratings. Nevertheless, attrition is not inevitable since not all individuals who are long-term residents in an L2-speaking environment end up with changes to their L1 accent. For example, de Leeuw et al. (2018a) showed that while one of their Albanian-English bilinguals completely neutralized the L1 phonemic contrast between light and dark laterals, and two additional ones did so only in coda position, others produced their laterals entirely like Albanian monolinguals. Similarly, in Mennen's (2004) study of tonal alignment, four out of five of her Dutch learners of Greek exhibited changes in their L1 alignment patterns, but one speaker did not, producing tonal alignment entirely natively in both languages (see also (de Leeuw et al. 2013; Major 1992)). Finally, instances of individual variation were found in studies of accent perception (Bergmann et al. 2016; de Leeuw et al. 2010). Thus, while 14 bilinguals in de Leeuw et al. (2010) received a clear non-native rating, 20 were consistently perceived as native.

1.2. L1 Use and Dual Language Activation in L1 Attrition

One of the variables that may account for such individual variation in L1 attrition of speech is language use. For example, Flege et al. (1997) showed that Italians in the United States had stronger foreign accents in L2 English if they used Italian a lot than if they used it rarely. Similarly, Lloyd-Smith et al. (2020) found a strong effect of Italian use scores on the perceived nativeness in Italian heritage speakers in Germany, while the age at which the heritage language was introduced was inconsequential. Stangen et al. (2015), in turn, found high non-native accents in the majority language German for Turkish heritage language speakers in Germany with high use of Turkish (see also (Kupisch et al. 2014)).

Similar effects of language use have also been documented in attrition contexts. Thus, Stoehr et al. (2017) examined VOT production in two groups of late Dutch-German bilinguals living in the Netherlands, L1 German speakers and L1 Dutch speakers. Native German speakers were exposed to their L1 only at home, whilst speaking Dutch in other environments, whereas the native Dutch speakers had more contact with their L1 given its status as the majority language, only coming into contact with L2 German at home. The study found that L2-immersed bilinguals produced nativelike L2 plosives, yet also exhibited L2-like characteristics in their L1 productions. Conversely, bilinguals living in the L1 environment did not produce nativelike L2 plosives but maintained nativelike L1 VOTs. Together, the results suggest that being immersed in an L2-speaking environment can be advantageous for L2 speech learning, but reduced L1 use may increase the likelihood of L1 attrition.

The idea that low L1 use should lead to attrition is based on the premise, consistent with exemplar theoretic and usage-based approaches, that language use reinforces memory representations, and that its absence may lead to retrieval difficulties (Bybee 2001). Nevertheless, the role of L1 use in attrition is not straightforward. First, a number of studies have shown that changes to L1 accents can occur despite continued high L1 use (Chang 2012; Mayr et al. 2012; Mennen 2004). For instance, Mayr et al. (2012), who investigated L1 attrition of speech in Dutch-English twin sisters, documented changes in L1 accent in the L2-immersed twin despite regular high use of her native Dutch. Mennen (2004), in turn, showed in her study of Dutch-Greek bilinguals in the Netherlands that L1 phonetic changes can even occur in an L1-speaking environment provided the frequency of L2 use is high. Second, L1 use and exposure must be seen as distinct from L2 immersion, in that residence in an L2-speaking environment can co-occur with wide and varied patterns of L1 communication. As such, simple measures of frequency and quantity of L1 contact may not be sufficient, since "[...] among bilinguals, L1 use does not necessarily equal L1 use" (Schmid 2007, p. 137). That is to say, L1 use encompasses a diverse range of situations that do not fit comfortably within a single definition, and therefore cannot be considered a single predictor of attrition.

One of these concerns situations that require co-activation of the L1 and L2. Thus, in de Leeuw et al.'s (2010) study, native German speakers in Anglophone Canada and the Netherlands were more likely to be perceived as foreign-accented in the L1 if they used German in contexts in which code-switching was likely to occur. Bilinguals who reported a high amount of L1 contact in situations with minimal expected code-switching, on the other hand, were less likely to be perceived as non-native, suggesting that L1 contact of this type may promote stability of pronunciation. Note, however, that in this study, participants were not directly asked whether they code-switched in specific settings. Rather, the authors postulated ex post facto that code-switching was more likely to occur in certain settings. These included L1 use with family members and friends in Canada and the Netherlands and use in church settings; in contrast, code-switching was deemed less likely to occur in work settings, during visits to Germany, and during telephone conversations and written correspondence with native German speakers.

These findings are consistent with a large body of evidence that has shown cross-linguistic interactions to occur in contexts of dual language activation, such as code-switching, where inhibition of the non-target language is particularly difficult (Green 1998). The state of activation of a bilingual's two languages at a given point in time is referred to as language mode (Grosjean 2001) and can range from bilingual mode, where both languages are fully activated, to monolingual mode, where the non-target language is inhibited as much as possible, although never entirely, based on sociolinguistic factors. Studies of phonetic code-switching have shown unidirectional interactions, in which the speech patterns of only one language are affected by those of the other one (e.g., (Muldner et al. 2019; Olson 2013)) as well as bidirectional interactions, in which both languages mutually affect each other's speech patterns (e.g., (Bullock and Toribio 2009; Piccinini and Arvaniti 2015)), with few studies revealing no effect of switching (but see (Grosjean and Miller 1994)).

1.3. The Present Study

The present study sought to build on previous work that has examined the role of L1 use and dual language activation in L1 attrition by investigating the perceived L1 accent of two groups of native Spanish speakers in the United Kingdom, (1) Spanish language teachers, and (2) non-teachers, alongside monolingual controls in Spain. As such, it is the first to examine L1 attrition of speech across specific professional groups. To the best of our knowledge, only one other study on L1 speech production has included individuals who teach their native language in an L2-speaking environment, that is, Chang (2019). However, unlike the present study, the speech of the L1 English speakers in that study, who taught their native language to L2 learners in Korea, was not compared to that of a group of non-teachers. Moreover, the focus of that study was the effect of bilinguals' L2 use on L1 pronunciation patterns.

The case of teachers is particularly pertinent, given the high proportion of foreign citizens who work teaching their native languages: Of an estimated 116,000 Spaniards in the United Kingdom between 2013 and 2015, nearly 10% were working in education (Office for National Statistics 2017).

While other migrants may also have frequent L1 contact, the experience of language teachers is quite distinct, given their high levels of L1 exposure and use under specific circumstances. Thus, language teaching is one of the few professions in which language is not merely a medium of communication, but also its object. As such, individuals who teach their native language to L2 learners may have what Chang (2019, p. 108) refers to as an "instructional orientation" towards the L1, which would typically encompass "high metalinguistic awareness and explicit knowledge of rules, norms, and standards" (ibid.). Moreover, the need for them to provide a clear, carefully articulated model for their students' pronunciation patterns means that they may be particularly concerned about retaining a native-like accent. Finally, teaching one's native language necessitates sustained high use of the L1. Together, these factors suggest that the L1 accent of individuals who teach their native language may be especially protected from attrition.

On the other hand, teaching one's L1 in an L2-speaking environment requires regular use of the L2, not only in social contexts but also professionally. Thus, even if foreign language teachers aim to maximize the use of the target language in the classroom, regular recourse to the ambient language, and the use of both languages in alternation, is virtually inevitable (Littlewood and Yu 2011; Turnbull and Dailey-O'Cain 2009). Moreover, recent pedagogical approaches, notably "translanguaging" (Cenoz and Gorter 2019), have moved away from strict adherence to monolingualism and actively embrace the use of more than one language in language classrooms, in line with Cook's (2008) notion of "multicompetence" (see also (Illman and Pietilä 2018)). Individuals teaching their native language, therefore, need to keep both their L1 and L2 fully activated for extended periods, and hence operate in a sustained bilingual language mode in the classroom (Grosjean 2001). As discussed previously, this has been shown to enhance the likelihood of cross-linguistic interactions, and as a result changes to individuals' L1 accent (cf. (de Leeuw et al. 2010)).

In addition, language teachers are regularly exposed to L1-influenced pronunciations in their students' L2 productions. However, the effect that foreign-accented input has on their native speech patterns is unclear. On the one hand, experimental studies examining phonetic convergence in native–non-native dyads have failed to document instances of native speaker accommodation towards the accents of non-native speakers (Kim 2009; Kim et al. 2011), suggesting that language teachers may be impervious to the influence of their students' accented speech patterns. On the other hand, in these studies, accommodation is based on singular events during which rapid phonetic adjustments are assessed in conversations with unfamiliar individuals, and hence they do not allow conclusions to be drawn about the effects of repeated exposure to, and interaction with, familiar foreign-accented speakers in professional educational settings. It is certainly plausible that sustained accented input of this kind may affect the representations of teachers' L1 speech sounds, in line with Chang's (2019) Incidental Input Hypothesis, which argues that ambient input is incidentally processed and cannot be ignored. Moreover, evidence from both adults who were raised in bilingual homes (Bosch and Ramon-Casas 2011) and bilingual children in immersion school settings (Caldas 2006; Mayr and Montanari 2015) supports the idea that foreign-accented input may affect L1 pronunciation patterns. Thus, Bosch and Ramon-Casas (2011) showed that Catalan-Spanish bilinguals who were raised with both languages and received inconsistent phonetic input produced Catalan/e-ɛ/ less accurately as adults than bilinguals raised in Catalan-only homes. Caldas (2006), in turn, reported that his daughters' L1 French was English-accented, which he attributed to their exposure to non-native speech at their dual language school in Louisiana. In contrast, his son, who was solely educated through the medium of English, but like his sisters received native French input in the home, had a native-like accent in French. Similarly, Mayr and Montanari (2015) found that the two Italian-English-Spanish trilingual children in their study had native-like VOT patterns in Spanish, but English-accented ones in Italian, even though both languages contain a prevoiced—short lag VOT contrast. The authors attributed this finding to the fact that the children were regularly exposed to English-accented Italian from their classmates in their Italian-English dual language school in Los Angeles, while they only learnt Spanish from their monolingual Mexican nanny.

Based on these considerations, the present study sought to answer three inter-related research questions. First, it aimed to find out whether Spanish speakers who teach their native language in an L2-speaking environment are perceived as more or less native-like in their L1 than non-teachers in the same L2 environment who rarely use it. Second, it sought to determine to what extent perceptions of non-nativeness are characterized by individual variation. Finally, it attempted to identify the specific accentual features that are associated with non-native speech in native Spanish teachers and non-teachers who are long-term residents in an L2-speaking environment.

2. Method

An accent rating experiment was carried out in which monolingual Spanish listeners, resident in Spain, were exposed to short extracts of Spanish speech from a picture-based narrative produced by two groups of native Spanish speakers in the United Kingdom, language teachers, and non-teachers, alongside monolingual controls in Spain. Listeners were asked to state whether they detected a non-native accent in the speech samples and to provide an indication of their level of confidence in their judgement. Moreover, if they considered a sample to sound non-native, they were prompted to identify the accentual features that had led them to this conclusion.

2.1. Participants

Two groups of consecutive bilingual Spaniards living in the United Kingdom were recruited to participate in the study: (1) Spanish language teachers (BIL-T, N = 10, 9 females), and (2) non-teachers (BIL-NT, N = 9, 5 females). Those in the latter group practise a diverse range of professions, ranging from social work to accountancy and nursing, and none habitually use Spanish in their communication at work or at home. The participants in BIL-T, in turn, were either employed as Spanish teachers in schools (N = 5) or in university settings (N = 5). Further to being long-term residents—that is, having lived continuously in the UK for at least five years—an inclusion criterion for both of these groups was that migration took place after the age of 18. In this way, any differences identified in their speech can be attributed to attrition as opposed to incomplete L1 acquisition (Schmid 2014).

In addition to the two bilingual groups, a group of monolingual Spaniards residing in Spain participated in the study (MON, N = 8, 7 females). The speakers in this group had never lived anywhere other than Spain, had never spoken a language other than Spanish at home, as a medium of education or at work, and reported low levels of proficiency in English or any other language. As such, they meet Best and Tyler's (2007) definition of functional monolinguals as "not actively learning or using an L2" (p. 16).

Participants were recruited through ELE-UK (www.eleuk.org) and the Instituto Cervantes (www.cervantes.es), both of which are institutions dedicated to the teaching of the Spanish language, and through Spanish departments at English universities as well as via existing networks in the United Kingdom and in Spain. They came from a range of regions in Spain with no systematic differences across the groups: Andalusia (BIL-T: 2, BIL-NT: 1, MON: 1), Asturias (BIL-NT: 1, MON: 1), Castile-La Mancha (BIL-NT: 1, MON: 1), Catalonia (BIL-T: 2, BIL-NT: 1, MON: 1), Galicia (BIL-T: 3), Murcia (BIL-T: 1), Madrid (BIL-T: 1, BIL-NT:1, MON: 1), Basque Country (BIL-T: 1), Valencia (BIL-NT: 4, MON: 4).

All subjects gave their informed consent for inclusion before they participated in the study. The research reported in this manuscript was reviewed and approved by the Cardiff School of Health SciencesResearch Ethics Committee, Cardiff Metropolitan University, United Kingdom (ethics reference number: UG-265).

Initial contact was established by email and, in order to ensure groups were matched for key variables, demographic and linguistic background information was collected by means of an online questionnaire created using Qualtrics XM (Qualtrics 2019). A summary of participant characteristics is included in Table 1. Comparisons on all variables in the table were made across the two bilingual groups, while comparisons across all three groups were only made on the first three variables in the table, that is, education, English proficiency, and chronological age, as well as on gender distributions.

Table 1. Participant characteristics.

	BIL-T (N = 10)		BIL-NT (N = 9)		MON (N = 8)	
	Median	**Min-Max**	**Median**	**Min-Max**	**Median**	**Min-Max**
Education	6.00	5.00–7.00	5.00	4.00–7.00	4.50	2.00–6.00
ENG proficiency	4.00	3.00–5.00	3.00	3.00–5.00	1.00	1.00–2.00
	Mean	**SE**	**Mean**	**SE**	**Mean**	**SE**
Chronological age (years)	41.60	3.11	33.56	2.16	31.63	1.29
AOA (years)	28.20	2.78	24.44	0.93	-	-
LOR (years)	13.10	2.18	8.89	1.84	-	-
Use of ENG at home	41.50%	12.26	79.67%	10.96	-	-
Use of SPAN at home	33.50%	12.03	9.22%	4.72	-	-
Use of ENG at work	54.20%	7.48	95.11%	2.77	-	-
Use of SPAN at work	40.60%	6.66	4.33%	2.69	-	-
Social use of ENG in UK	63.30%	5.72	76.22%	6.96	-	-
Social use of SPAN in UK	33.00%	5.16	23.78%	6.96	-	-

Notes: AOA = age of arrival in the UK; LOR = length of residence.

2.1.1. Comparisons across the Two Bilingual Groups

The two bilingual groups were carefully matched on a range of background variables[1]. Thus, they did not differ from each other in gender distribution (Chi-square test: $\chi^2(1) = 2.898$, $p = 0.089$), chronological age (BIL-T (mean: 41.60, SE: 3.11); BIL-NT (mean: 33.56, SE: 2.16); Independent t-test: $t(17) = 2.08$, $p = 0.053$), age of arrival in the UK (BIL-T (mean: 28.20, SE: 2.78); BIL-NT (mean: 24.44, SE: 0.93); Independent t-test: $t(17) = 1.223$, $p = 0.238$) or length of residence (BIL-T (mean: 13.10, SE: 2.18); BIL-NT (mean: 8.89, SE: 1.84); Independent t-test: $t(17) = 1.458$, $p = 0.163$). Moreover, they were matched in terms of their highest level of education (BIL-T (median: 6.00, min-max: 5.00–7.00); BIL-NT (median: 5.00, min-max: 4.00–7.00); Mann–Whitney test: $U = 27.500$, $p = 0.126$), using a seven-point Likert scale ranging from 1 (less than secondary school education) to 7 (doctorate), as well as their self-reported competence in English (BIL-T (median: 4.00, min-max: 3.00–5.00); BIL-NT (median: 3.00, min-max: 3.00–5.00); Mann–Whitney test: $U = 35.000$, $p = 0.374$), based on a six-point Likert-type scale ranging from 1 (less than basic knowledge of English) to 6 (Native or near-native proficiency) in line with the classifications of the Common European Framework of Reference for Languages (Council of Europe 2001).

The bilingual groups were also matched on some of their language use patterns. Thus, they did not differ in their estimated use of Spanish and English in social situations outside their home and work in the UK (Spanish: BIL-T (mean: 33.00, SE: 5.16); BIL-NT (mean: 23.78, SE: 6.96); Independent t-test: $t(17) = 1.079$, $p = 0.296$; English: BIL-T (mean: 63.30, SE: 5.72); BIL-NT (mean: 76.22, SE: 6.96); Independent t-test: $t(17) = 1.446$, $p = 0.166$), the amount of time they spent in Spain per year (BIL-T (median: 1.00 (<1 month), min-max: 1.00 (<1 month) to 2.00 (1–3 months)), BIL-NT (median: 1.00 (<1 month), min-max: 1.00 (<1 month) to 2.00 (1–3 months)); Mann–Whitney test: $U = 43.500$, $p = 0.879$), the frequency of spoken contact with family and friends in Spain, for example, via telephone conversations (BIL-T (median: 1.00 (once or twice a week), min-max: 1.00 (once or twice a week) to 3.00 (less than once a month)); BIL-NT (median: 1.00 (once or twice a week), min-max: 1.00 (one or twice a week) to 3.00 (less than once a month)); Mann–Whitney test: $U = 38.500$, $p = 0.492$), or the frequency of written contact with family and friends in Spain, for example, email correspondence

[1] To compare groups on scalar variables, such as chronological age, we ran parametric tests (independent t-test; one-way ANOVAs); for comparisons on ordinal variables and Likert scales, we ran non-parametric tests (Mann–Whitney test; Kruskal–Wallis test); the relation between nominal variables, in turn, was explored using chi-squared tests. When running independent samples t-tests across the two bilingual groups on the use of English and Spanish at work as well as on the use of Spanish at home, the variances turned out not be equal based on Levene's tests. In these cases, the t-values, p-values, and degrees of freedom were adjusted accordingly.

(BIL-T (median: 2.00 (once or twice a day), min-max: 1.00 (multiple times a day) to 4.00 (once or twice a month)); BIL-NT (median: 2.00 (once or twice a day), min-max: 1.00 (multiple times a day) to 3.00 (once or twice a week)); Mann–Whitney test: $U = 27.500$, $p = 0.129$).

In contrast, crucially, the two groups differed from each other in terms of their language use patterns in work, and to a lesser extent at home. Thus, the BIL-T group used English significantly *less* at work (BIL-T (mean: 54.20, SE: 7.48), BIL-NT (mean: 95.11, SE: 2.77); Independent t-test: $t(11.390) = 5.130$, $p < 0.0005$) and at home (BIL-T (mean: 41.50, SE: 12.26), BIL-NT (mean: 79.67, SE: 10.96); Independent t-test: $t(17) = 2.30$, $p = 0.034$) than the BIL-NT group, but Spanish significantly more at work (BIL-T (mean: 40.60, SE: 6.66), BIL-NT (mean: 4.33, SE: 2.69); Independent t-test: $t(11.862) = 5.047$, $p < 0.0005$) than the BIL-NT group. On the other hand, the two groups did not differ significantly from each other in their use of Spanish at home (BIL-T (mean: 33.50, SE: 12.03), BIL-NT (mean: 9.22, SE: 4.72); Independent t-test: ($t(11.674) = 1.878$, $p = 0.086$). Note that two of the BIL-T speakers and one of the BIL-NT speakers live by themselves and therefore indicated no use of any language in the home. Note also that the BIL-T speakers, but not the BIL-NT speakers, indicated occasionally using a language other than Spanish or English that was not specified further. This accounted for circa 5% of the use patterns at work and 7% at home.

2.1.2. Comparisons across the Monolingual Group and the Two Bilinguals Groups

Finally, comparisons were made across all three groups, that is, BIL-T, BIL-NT, and MON. They differed significantly on self-rated competence in English (MON (median: 1.00, min-max: 1.00–2.00), BIL-T (median: 4.00, min-max: 3.00–5.00); BIL-NT (median: 3.00, min-max: 3.00–5.00); Kruskal–Wallis test: $\chi^2(2) = 10.16$, $p = 0.006$) with a Dunn's post-hoc test revealing significantly lower scores for the MON group than BIL-T ($p = 0.002$) and BIL-NT ($p = 0.016$). Moreover, while the MON speakers did not differ from the other two groups in terms of gender distribution (Chi-squared test: $\chi^2(2) = 3.873$, $p = 0.144$), they differed in chronological age (MON (mean: 31.63, SE: 1.29), BIL-T (mean: 41.60, SE: 3.11); BIL-NT (mean: 33.56, SE: 2.16); One-way ANOVA: $F(2,24) = 4.810$, $p = 0.018$) and formal education level (MON (median: 4.50, min-max: 2.00–6.00), BIL-T (median: 6.00, min-max: 5.00–7.00); BIL-NT (median: 5.00, min-max: 4.00–7.00); Kruskal–Wallis test: $\chi^2(2) = 6.74$, $p = 0.034$), with the MON group significantly younger ($p = 0.030$) and less well educated ($p = 0.029$) than the BIL-T group, but not the BIL-NT group ($p > 0.05$).

2.2. Speech Materials

Participants audio-recorded themselves telling the story "I will help you" (Abbott et al. 2015) in Spanish. To do this, they were given access to an adapted version of the picture book online, which contained 17 pictures, but with all words removed. Participants could view the pictures as many times as they wished to ensure they understood the story before recording. Recordings were completed with a mobile phone or computer in a quiet environment, avoiding background noise, to promote optimum quality for subsequent use in the accent rating experiment. They were asked not to plan the exact wording beforehand and to imagine telling the story to a monolingual Spanish child. This approach was chosen to obtain quasi-spontaneous speech, whilst ensuring comparable samples in terms of lexical and grammatical content, and thus minimizing the likelihood of judgements resulting from differences in linguistic complexity (Schmid and Hopp 2014).

From each of the 27 narratives, a randomly selected speech sample of approximately 15 s was extracted in PRAAT (Boersma and Weenink 2019). This duration was considered sufficient for listeners to make a reliable judgement (de Leeuw et al. 2010; Flege 1984; Schmid and Hopp 2014). In order to minimize the likelihood that the listeners' judgements are based on areas other than pronunciation, samples were carefully screened to ensure they contained no lexical or grammatical errors and constituted grammatically complete utterances. Long pauses and hesitations were also avoided. A one-way ANOVA revealed no statistically significant difference in sample duration across groups (Mean BIL-T: 16.33 (SD: 1.14); Mean BIL-NT: 15.53 (SD: 0.763); Mean MON: 15.89 (SD: 0.524);

$F(2,24) = 1.979$, $p = 0.160$), nor in terms of speaking rate, as measured in syllables per second (Mean BIL-T: 5.21 (SD: 0.649); Mean BIL-NT: 5.54 (SD: 0.852); Mean MON: 5.70 (SD: 0.502); $F(2,24) = 1.201$, $p = 0.318$). To further reduce variability across samples, peak intensity was normalized, using PRAAT software (Boersma and Weenink 2019).

2.3. Listeners

The samples were presented as part of an online questionnaire created in Qualtrics XM software (Qualtrics 2019), which was distributed via an anonymous link to students at the Faculty of Education, University of A Coruña as well as existing networks across Spain. A total of 28 native Spanish listeners (20 females) with a mean age of 32 (SD: 11.25) completed the online accent rating experiment. Competence in English was controlled for with none of the listeners reporting higher than intermediate proficiency (mean 2.5, SD 0.75) comparable to the MON speakers' scores (cf. Table 1). Like the MON speakers, the listeners had never lived outside Spain and had never spoken a language other than Spanish at home, as a medium of education, or at work. Like the speakers, they come from a variety of regions, including Andalusia ($N = 6$), Castile-La Mancha ($N = 1$), Catalonia ($N = 3$), Extremadura ($N = 1$), Galicia ($N = 3$), Madrid ($N = 1$) and Valencia ($N = 13$).

2.4. Experimental Procedure

As the experiment was conducted online, listeners were given detailed written instructions regarding the task at hand. They were asked to use headphones, and an audio test was incorporated into the questionnaire to ensure adequate browser and volume settings had been selected. Participants were informed they would hear samples from fluent Spanish speakers, though no indication of whether they were native or not was given. Following the method established by Moyer (1999) and adopted in various studies on bilingual populations since then (e.g., (Bergmann et al. 2016; de Leeuw et al. 2010; Lloyd-Smith et al. 2020)), samples were played in random order and after each recording listeners were instructed to give a binary rating of the speaker's accent (native/non-native), indicating subsequently their degree of confidence (confident/neither confident nor not confident/not confident). They were further instructed to select "non-native" in the event they detected a non-native accent, however slight. Listeners heard each sample only once and were asked to guess if unsure, indicating their lack of confidence accordingly.

For samples rated "non-native", a follow-up question was included immediately after the rating was given, requesting details of what aspects of pronunciation had created a perception of non-nativelikeness, as well as any specific words that sounded non-native. In addition to the rating task, the questionnaire contained a range of demographic and language background questions to ensure the listeners met the inclusion criteria.

No time limit was imposed for responding and listeners controlled the pace at which they progressed through the samples. They were encouraged to take as many breaks as they deemed necessary. The average duration for the experiment was 25 min.

2.5. Analysis

In line with previous accent rating experiments (Bergmann et al. 2016; de Leeuw et al. 2010; Moyer 1999), listeners' responses were converted to a six-point scale in which a "native" rating marked as "confident" appeared at one end of the scale (1) and a "confident" rating as "non-native" at the other (6). As such, the lower the numerical foreign accent rating (FAR), the nearer to nativelike the speaker was perceived to be. The experimental data were subsequently transferred to a CSV file for statistical analysis. In order to assess whether the groups differ in their FAR, linear mixed-effects models were run in R (R Core Team 2018) using the LmerTest function (Kuznetsova et al. 2017). To analyze the features identified by the listeners, content analysis was used (Krippendorff 2018). This first involved screening responses for relevant phonetic information. Comments that did not relate to accentual features were disregarded. Items referring to accentual features, in turn, were initially coded as relating to either segmental or suprasegmental phenomena before being assigned to more specific subcategories. These were then quantified. As a measure of reliability, coding was repeated on all 174 comments, yielding an agreement score of 95.98%. Divergences between the two sets of analysis only concerned a small number of comments with unclear/ambiguous meanings. For example, reference to "una pronunciación muy marcada, muy fuerte" (a very marked pronunciation, very strong) was coded as referring to rhythm/stress in the first analysis, but as being too general to include in the re-analysis. These comments were discarded from further analysis.

3. Results

3.1. Accent Rating

To assess inter-rater reliability, we ran a Cronbach's alpha analysis across the ratings made by the 28 listeners. The results revealed a value of 0.81, which suggests a high degree of homogeneity. Figure 1 depicts the distribution of FAR scores across the three groups.

Inspection of the figure suggests that the samples were predominantly perceived to be native-like, with median scores of "1" for the participants in BIL-NT and MON, and of "2" for the participants in BIL-T, although the scores in all groups exhibited a certain degree of variation. Overall, a total of 221 of the $28 \times 27 = 756$ samples were rated as non-native, that is, 29.23%, with 107 (i.e., 14.15%) attracting the highest FAR score of "6", that is, "non-native with certainty". To examine whether the FAR scores differed across the groups, linear mixed-effects models were run in R (R Core Team 2018), with "group" as fixed effect and "participant" as random intercept. Using the LmerTest function (Kuznetsova et al. 2017), the Satterthwaite approximation was used to obtain degrees of freedom, from which *p*-values could be calculated.

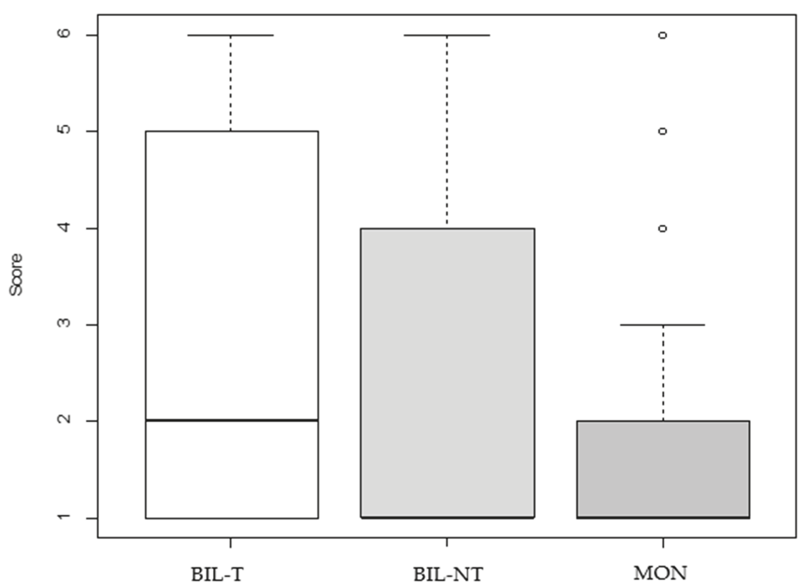


Figure 1. Distribution of FAR scores by group.

Our initial model was run on all 756 ratings and across the three groups. The results, depicted in Table 2, revealed highly significant between-group differences ($p < 0.001$). This analysis was subsequently followed up with pairwise comparisons, with a Bonferroni-adjusted α-level of 0.0167.

Table 2. Results of linear mixed-effects models: FARs.

Model		β	SE	t	p
All groups	Intercept	3.59455	0.20179	17.813	<0.001
	Group	−0.54772	0.07986	−6.859	<0.001
BIL-T vs. BIL-NT	Intercept	3.8944	0.2751	14.16	<0.001
	Group	−0.7766	0.1625	−4.78	<0.001
BIL-T vs. MON	Intercept	3.6567	0.2115	17.29	<0.001
	Group	−0.53884	0.08054	−6.69	<0.001
BIL-NT vs. MON	Intercept	2.9435	0.4012	7.337	<0.001
	Group	−0.3011	0.1552	−1.94	0.053

The results revealed significantly higher and thus less native-like FAR scores for the participants in BIL-T than in BIL-NT ($p < 0.001$) and MON ($p < 0.001$). The difference between the latter two, in contrast, was not significant ($p = 0.053$). Together, these results suggest that the L1 accent of Spaniards in non-teaching professions in the UK was perceived as equally native-like as that of monolinguals resident in Spain. Spaniards teaching their L1 in educational settings in the UK, in contrast, whilst also attracting relatively low FAR scores, were perceived as significantly less native-like, suggesting a certain degree of L1 attrition.

3.2. Perceived Non-Native Features

All 28 listeners provided comments on the samples they deemed non-native; however, this was only the case for 174 of the 221 samples (i.e., 78.73%), while 47 of the non-native ratings were left uncommented. Following a careful screening, 71 of the 174 comments were removed from the analysis as they were too general, referring, for example, just to "pronunciation of some words" or "the speaker's accent", and an additional three were removed that referred to features unrelated to pronunciation, for example, lexical or grammatical choice. The remaining 100 comments were analysed further; of these, 84 referred to a single feature, while 13 referred to 2 features, and 2 to 3 features, for a total of 116 feature tokens. Table 3 shows a breakdown of the features identified, alongside illustrative examples. Since they did not exhibit any systematic differences between the speakers in BIL-T and BIL-NT, the data were pooled.

Table 3. Perceived non-native features.

Features Identified	Tokens (N = 116)	%	Example
Segmental features	75	64.7%	
Consonants	63	54.3%	
s	28	24.1%	*La pronunciación de la S muy forzada* (Very forced pronunciation of the S)
r	24	20.7%	*La pronunciación de la letra R en algunas palabras* (Pronunciation of the letter R in some words)
t	4	3.4%	*Pronunciación de las Tes* (Pronunciation of the Ts)
d	3	2.6%	*Pronunciación letra D* (Pronunciation of the letter D)

Table 3. *Cont.*

Features Identified	Tokens (*N* = 116)	%	Example
β	1	<1%	*Pronunciación letra V* (Pronunciation of the letter V)
l	1	<1%	*Dudosa pronunciación con la letra L* (Dubious pronunciation of the letter L)
ʎ	1	<1%	*La "ll" de polluelos ha sonado rara* (The "ll" in "polluelos" (chicks) sounded odd)
General	1	<1%	*Parece hablar con una pronunciación no nativa en distintas consonantes* (She seems to speak with a non-native pronunciation in different consonants)
Vowels	8	6.9%	
General	3	2.6%	*Ha abierto demasiado las vocales.* (She opened her vowels too much)
Diphthongs	2	1.7%	*La pronunciación del último diptongo* (Pronunciation of the final diphthong)
a	1	<1%	*La pronunciación de las As* (Pronunciation of the As)
e	1	<1%	*La pronunciación de la E de forma más cerrada* (Pronunciation of E [is] closer)
i	1	<1%	*Las "y" muy señalada* ("y" (and) [were] very marked)
Phoneme omission	4	3.4%	*No pronuncia todos los fonemas* (She does not pronounce all phonemes)
Suprasegmental features	41	35.3%	
Intonation	26	22.4%	*Me parece una persona no nativa por la musicalidad en la pronunciación, más típico del italiano* (He seems non-native to me due to the musicality in the pronunciation; more typical of Italian)
Rhythm/Stress	8	6.9%	*Acentuación en la terminación de palabras* (Accentuation/stress at the end of words)
Speaking rate	7	6.0%	*La aceleración al hablar* (Acceleration when speaking)

Inspection of the table shows that judgements of non-nativeness were based on both segmental and suprasegmental features, albeit with a preponderance of the former. Amongst segments, listeners most commonly perceived consonantal items as non-native, notably realizations of /s/ and rhotic consonants, but some also referred to vowel deviations and phoneme omissions. Comments on suprasegmental items predominantly referred to intonation, mostly expressed in terms of "melodía" (melody) or "musicalidad" (musicality), but there were also some mentions of rhythm/stress and speaking rate.

3.3. Individual Variation

Finally, in addition to the analysis at the group level, we investigated individual variation. This was done by converting median FARs into a categorical rating of "clearly native" (between 1.0 and 2.5), "uncertain" (greater than 2.5 but less than 4.5), and "clearly non-native" (between 4.5 and 6.0) following de Leeuw et al.'s (2010) approach. The categorizations for the participants in the three groups are shown in Table 4.

Table 4. Categorization of nativeness by group.

	BIL-T (N = 10)	BIL-NT (N = 9)	MON (N = 8)
Clearly native	5 (50%)	7 (77.8%)	8 (100%)
Uncertain	1 (10%)	1 (11.1%)	-
Clearly non-native	4 (40%)	1 (11.1%)	-

Inspection of the table shows that, as one would expect, all MON speakers were consistently classed as "clearly native". In contrast, in line with previous work on attrition (e.g., (de Leeuw et al. 2010, 2018a; Mennen 2004)), the results for the two bilingual groups were more varied. Thus, although the BIL-NT speakers were not found to differ from the MON ones at the group level, as we have seen, the analysis of individual classifications shows that one BIL-NT speaker was considered "uncertain" and another one "clearly non-native". At the same time, while 4 in the BIL-T group were classed as "clearly non-native" and one as "uncertain", half of them were considered "clearly native". As a result, teaching one's native language in an L2-speaking environment does not automatically lead to perceived attrition in L1 speech; it merely appears to increase its likelihood. Table 5 displays the characteristics of the participants identified as non-native.

Table 5. Characteristics of participants perceived as non-native.

Participant	Gender	Age	Region	EN Proficiency	AOA (Years)	LOR (Years)	Median FAR	Non-Native Ratings	Features Identified
BIL-T_1	F	33	GA	5	25	7	5	20	intonation (3) s (2) r (1) e (1)
BIL-T_2	F	27	CT	4	23	5	5	21	s (3) t (2) intonation (2)
BIL-T_3	F	48	GA	5	34	14	5	21	r (6) s (4) intonation (1) l (1) d (1) rhythm/stress (1)
BIL-T_4	F	41	PV	6	23	18	5	20	intonation (5) r (4) s (3) rhythm/stress (2) t (1) vowels (1)
BIL-NT_1	F	34	AN	5	27	6	4.5	17	r (2) phoneme omission (1) intonation (1) rhythm/stress (1) consonants (1) d (1) s (1)

Note: AN = Andalusia; CT = Catalonia; GA = Galicia; PV = Basque Country; the figures in parenthesis denote the number of comments per feature.

As the table shows, all participants considered "clearly non-native" were female, aged between 27 and 48 years, and considered their English competence as upper intermediate to near-native. They had moved to the United Kingdom in their twenties or thirties and had been living there between 5 and 18 years. Of the 28 listeners, 20 or more considered the four BIL-T speakers as non-native; slightly fewer listeners, that is, 17, classified BIL-NT_1 as non-native. The latter also received a slightly lower FAR and was hence perceived as less clearly non-native than the four BIL-T speakers. Finally, the table shows that the Spanish accent of each of these participants was associated with multiple non-native features. All were perceived to produce their L1 with non-native intonation patterns and realizations of /s/, and all but one, that is, BIL-T_2, were perceived to realize Spanish rhotics non-natively.

4. Discussion

This study aimed to gain a better understanding of the role of L1 use and dual language activation in the perceived attrition of native speech patterns. To this end, we examined the L1 Spanish accent of two groups of native Spanish speakers who are long-term residents in the UK, Spanish language teachers and non-teachers, alongside monolingual Spanish speakers in Spain in an accent perception experiment. The results revealed significantly greater non-native ratings for the teachers than the non-teachers and the monolinguals, but no difference between the latter two, with listeners' impressions of non-nativeness based on a range of segmental and suprasegmental features. An analysis of individual patterns, in turn, showed a fair amount of variation, with half of the speakers in BIL-T perceived as "clearly native" and one of the BIL-NT speakers as "clearly non-native". In what follows, the implications of these findings will be discussed.

To begin with, let us consider why the participants in BIL-T were perceived as significantly more foreign-accented in their L1 than monolinguals in Spain. At first glance, this finding is surprising. After all, they regularly use their L1 both in work and outside of it, and Spanish plays an essential role in their professional identity. As Chang (2019, p. 108) states, being a language teacher typically comes with an "instructional orientation", and likely coincides with a particular concern for retaining native-like proficiency in the L1, including its accent, although this was not formally assessed here. One might expect these factors to provide a certain degree of protection from attrition. However, this was not the case in the present study, at least not at the group level.

The likely reason for the perceived attrition in the teachers' L1 accent is dual language activation, which, in turn, is a direct consequence of the specific professional setting in which they operate. In other words, it is essentially impossible for foreign language teachers who teach their L1 in an L2-speaking environment to activate only their L1 during classroom activities and only their L2 outside of it, and hence function in alternate monolingual language modes (Grosjean 2001). Instead, both their languages need to be highly active for most or all of the time, resulting in them operating in a sustainable bilingual language mode. This will be true even if the extent of dual activation varies somewhat from context to context. For example, it is likely to be particularly high during activities that actively encourage a bilingual approach, such as translanguaging (Cenoz and Gorter 2019), while it will be comparatively lower during activities in which sole use of the target language is encouraged, in particular in students with high L2 proficiency levels. Nevertheless, whatever the specific circumstances, the very nature of foreign language classroom settings makes dual activation inevitable.

Crucially, dual activation has been shown to lead to cross-linguistic interactions in speech patterns. Such interactions have been widely attested in contexts of phonetic code-switching (Amengual 2018; Bullock and Toribio 2009; Muldner et al. 2019; Piccinini and Arvaniti 2015), where cognitive demands to inhibit the non-target language are particularly high. While they may initially occur in such circumstances, that is, during ad hoc dual language activation, over time they may give rise to more persistent accentual changes and become entrenched. This is likely to have happened to the teachers in the present study and is consistent with de Leeuw et al.'s (2010) finding that L1 attrition was more common in native German speakers in Anglophone Canada and the Netherlands who regularly used their L1 in contexts of code-switching than those who did not.

In addition, unlike the non-teachers, the teachers will have been systematically exposed to non-native Spanish accents via their students' productions. These may have either independently caused the observed changes in their L1 accent or enhanced the effects of their own concurrent use of the two languages, thereby reinforcing deviations from monolingual Spanish patterns. While the direct effect of sustained English-accented input in Spanish cannot be isolated in the present context, it will have led to an additional burden on teachers' inhibitory control mechanisms. The suggestion that foreign-accented input can increase the likelihood of non-native speech patterns is certainly consistent with evidence from adults raised in bilingual homes (Bosch and Ramon-Casas 2011) as well as bilingual and multilingual children in immersion school settings (Caldas 2006; Mayr and Montanari 2015), although its role in L1 attrition of speech needs to be explored further in future research. Taken together, the results for the participants in BIL-T suggest that, ironically, it is the very nature of the professional context in which teachers operate, with its requirement to keep both languages active and the need to switch between them, that enhances the likelihood of L1 attrition.

The participants in BIL-NT, in contrast, do not face these cognitive demands in a professional setting. While they work in a diverse range of areas, such as nursing, social work, and accountancy, none of them involve professional use of Spanish. As a result, the BIL-NT speakers virtually exclusively use their L2 in work, and hence operate in a consistent monolingual English language mode. Their lower overall amount of L1 use (and greater amount of L2 use), compared with the BIL-T group, in turn, did not lead to perceived attrition since they were rated the same as monolingual controls in Spain. Previous research suggests a somewhat ambiguous role for overall amount of language use in L1 attrition: while some studies have shown an effect of reduced L1 contact on attrition of speech patterns (e.g., (Stoehr et al. 2017)), others either revealed no effect (e.g., (Hopp and Schmid 2013)), or exhibited mixed results. For example, Chang (2019) showed no greater overall persistence in L1 phonetic drift in English-Korean bilinguals with high L2 use compared to those with low L2 use—only one of three areas investigated yielded a significant effect. While it is conceivable that the complete lack of L1 use over many years may cause attrition, independent of other factors, due to the gradual loss of long-term memory representations, this was not the case here. After all, even though the participants in BIL-NT hardly ever used their L1 in work contexts, they indicated using it regularly in social interactions outside of work as well as in written and spoken forms of remote communication with family and friends in Spain. The reduction in L1 use that typically occurs in L2 immersion contexts is hence unlikely to cause L1 attrition of speech in and of itself. It appears that what is critical is the contexts in which the L1 is used (cf. (Schmid 2007)). In the present study, it may well be the absenc*e* of L1 use in the kinds of contexts in which the teachers use their native language professionally, that has protected the BIL-NT speakers' speech from attriting. At the same time, their high L2 competence will have protected them from experiencing L1 phonetic drift as a result of a novelty effect (Chang 2012, 2013).

These considerations notwithstanding, the results of the present study also show a fair amount of individual variation, with half of the participants in BIL-T being perceived as "clearly native-like" and one participant in BIL-NT as "clearly non-native". Moreover, while the BIL-T participants were rated as significantly more non-native than those in BIL-NT and MON, their median FAR was "2", that is, "native-like with medium confidence". This suggests that L1 attrition in the context of teaching one's L1 in an L2-speaking environment is by no means inevitable. Perhaps the five teachers in BIL-T who were rated as "clearly native" in Spanish developed enhanced inhibitory control which allowed them to counteract cross-linguistic interactions from dual language activation and exposure to foreign-accented speech by their students. This may have coincided with a range of factors relating to individual differences, such as attitudinal, socio-psychological, and cognitive ones. For example, they may ascribe particular importance to the retention of a native accent in Spanish. Or they may have a particular phonetic talent (e.g., (Jilka 2009; Lewandowski and Jilka 2019)). Moreover, they may actually be perceived as non-native, but only in settings not assessed here, for example, in casual encounters (Major 1992). By the same token, the absence of particular skills or attitudes may explain attrition in BIL-NT_1's L1 accent. However, explanations of this nature remain wholly speculative as these variables were not investigated in the present study. Suffice it to say that L1 attrition of speech is a complex multi-factorial phenomenon (cf. (Kartushina et al. 2016b)) and that the patterns observed here must have been caused, in part, by factors other than dual language activation and language use. Although challenging, future work, based on a larger sample of potential attriters, is needed that systematically teases the various predictor variables for L1 attrition of speech apart and includes a more sophisticated approach to the assessment of L1 use. In the context of language teachers, this could involve obtaining details on interaction patterns with students of varying levels of proficiency during different types of classroom activity, but also language use and code-switching patterns with fellow foreign language teachers outside the classroom.

While we have so far discussed differences between the teachers' and non-teachers' use of languages at work, their language use patterns at home also need to be considered. Our results showed that BIL-T not only differed from BIL-NT participants in their language patterns in the workplace, but also in their language patterns at home. Crucially though, the language differences at home only pertained to the use of the L2, which was used more frequently by the non-teachers than the teachers. In contrast, no differences were found between the two groups in their use of Spanish at home. This shows that the perceived attrition in the teachers' L1 accent, cannot be explained by a reduction in L1 use at home, given that their amount of L1 use was similar to that of the non-teachers.

Finally, let us consider the features that the listeners associated with non-native speech. They encompass a range of consonants, vowels, and prosodic phenomena, in particular realizations of /s/, rhotics and intonation patterns, in line with evidence that perceptions of non-nativeness arise from the interplay between segmental and suprasegmental characteristics (Ulbrich and Mennen 2016). Importantly, there were no systematic differences in the features associated with non-nativeness in the BIL-T and BIL-NT speakers. Moreover, the speech of all speakers who were identified as "clearly non-native" was characterized by multiple non-native features and at both segmental and suprasegmental levels. This suggests that listeners did not erroneously mistake them as non-native due to their unfamiliarity with individual features that are associated with native dialectal variation, such as the phenomenon of seseo/ceceo in the context of /s/ (Martínez-Celdrán et al. 2003). While the features identified must have been perceptually salient for the listeners, their relative importance to the impression of non-nativeness remains unclear. Moreover, the listeners' judgements may have been influenced by accentual patterns that they were not consciously aware of or that they were unable to verbalize. It is also difficult to ascribe the features to specific types of interaction with L2 English, for example, assimilation or dissimilation patterns (cf. SLM (Flege 1995; Flege and Bohn 2020)), due to a lack of detail in the comments provided. Future research exploring the salience of features in global accent ratings is needed that extends the work presented here, using a more sophisticated methodology, such as an interactive interview-based approach (Mayr et al. 2020) or one that allows listeners' judgements to be linked directly to specific items in the speech samples (Montgomery and Moore 2018).

5. Conclusions

The present study examined the role of L1 use and dual language activation in L1 attrition by investigating the perceived L1 accent of two groups of native Spanish speakers in the United Kingdom: (1) Spanish language teachers, who use their L1 regularly in professional settings that require frequent switching between Spanish and English, and (2) non-teachers, who virtually never use their L1 in the workplace. In addition, the study included a control group of monolingual speakers in Spain. As such, this study is the first to examine L1 attrition of speech systematically in a specific professional group. The results of a global accent rating experiment revealed significantly greater non-native ratings for the teachers than the non-teachers and the monolingual controls, but no difference between the latter two. Listeners' impressions of non-nativeness, in turn, were based on a range of segmental and suprasegmental features, notably /s/, rhotics and intonation. These results suggest that language teachers who teach their L1 in an L2-speaking environment may be particularly prone to L1 attrition. This is likely due to a need to co-activate both their languages in professional settings as well as regular exposure to non-native speech from L2 learners. In contrast, low L1 use was not associated with non-native features in the non-teachers' Spanish accents. Together, the findings hence suggest that cross-linguistic interaction is more likely to lead to L1 attrition of speech than reduced L1 use in and of itself. However, since not all teachers were perceived as non-native, future research based on a larger sample is needed that assesses the factors further that facilitate or hinder L1 attrition in such educational settings.

Author Contributions: Conceptualization, R.M., D.S. and I.M.; methodology, R.M. and D.S.; formal analysis, R.M. and D.S.; writing—original draft preparation, R.M. and D.S.; writing—review and editing, R.M., D.S. and I.M.; visualization, D.S. and R.M.; supervision, R.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

Abbott, Andrea, Olivia Villet, and Fathima Kathrada. 2015. I Will Help You. Available online: http://bookdash.org/i-will-help-you-by-fathima-kathrada-olivia-villet-and-andrea-abbott/ (accessed on 7 October 2018).

Ahn, Sunyoung, Charles B. Chang, Robert DeKeyser, and Sunyoung Lee-Ellis. 2017. Age Effects in First Language Attrition: Speech Perception by Korean-English Bilinguals. *Language Learning* 67: 694–733. [CrossRef]

Amengual, Mark. 2018. Asymmetrical interlingual influence in the production of Spanish and English laterals as a result of competing activation in bilingual language processing. *Journal of Phonetics* 69: 12–28. [CrossRef]

Bergmann, Christopher, Amber Nota, Simone A. Sprenger, and Monika S. Schmid. 2016. L2 immersion causes non-native-like L1 pronunciation in German attriters. *Journal of Phonetics* 58: 71–86. [CrossRef]

Best, Catherine T., and Michael D. Tyler. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In *Language Experience in Second Language Speech Learning. In Honor of James Emil Flege*. Edited by Ocke-Schwen Bohn and Murray J. Munro. Amsterdam: John Benjamins, pp. 13–34.

Boersma, Paul, and David Weenink. 2019. PRAAT: Doing Phonetics by Computer (Version 6.0.46). Computer Program. Available online: http://www.fon.hum.uva.nl/praat/ (accessed on 18 January 2019).

Bosch, Laura, and Marta Ramon-Casas. 2011. Variability in vowel production by bilingual speakers: Can input properties hinder the early stabilization of contrastive categories? *Journal of Phonetics* 39: 514–26. [CrossRef]

Bullock, Barbara E., and Almedia J. Toribio. 2009. Trying to hit a moving target: On the sociophonetics of code–switching. In *Interdisciplinary Approaches to Code-Switching*. Edited by Ludmila Isurin, Donald Winford and Kees de Bot. Amsterdam: John Benjamins, pp. 189–206.

Bybee, Joan. 2001. *Phonology and Language Use*. Cambridge: Cambridge University Press.

Bylund, Emanuel, Kenneth Hyltenstam, and Niclas Abrahamsson. 2013. Age of acquisition effects or effects of bilingualism in second language ultimate attainment'. In *Sensitive Periods, Language Aptitude, and Ultimate L2 Attainment*. Edited by Gisela Grañena and Mike Long. Amsterdam: John Benjamins, pp. 69–101.

Caldas, Stephen J. 2006. *Raising Bilingual Biliterate Children in Monolingual Cultures*. Clevedon: Multilingual Matters.

Cenoz, Jasone, and Durk Gorter. 2019. Multilingualism, translanguaging, and minority languages in SLA. *The Modern Language Journal* 103: 130–35. [CrossRef]

Chang, Charles. B. 2012. Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics* 40: 249–68. [CrossRef]

Chang, Charles B. 2013. A novelty effect in phonetic drift of the native language. *Journal of Phonetics* 41: 520–33. [CrossRef]

Chang, Charles B. 2019. Language change and linguistic inquiry in a world of multicompetence: Sustained phonetic drift and its implications for behavioral linguistic research. *Journal of Phonetics* 74: 96–113. [CrossRef]

Cho, Mi-Hui, and Shinsook Lee. 2016. The impact of different L1 and L2 learning experience in the acquisition of L1 phonological processes. *Language Sciences* 56: 30–44. [CrossRef]

Cook, Vivian. 2008. *Second Language Learning and Language Teaching*, 4th ed. London: Hodder Education.

Council of Europe. 2001. Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Available online: https://rm.coe.int/1680459f97 (accessed on 20 August 2020).

de Bot, Kess, Wander Lowie, and Marjolijn Verspoor. 2007. *A Dynamic View as a Complementary Perspective*. Bilingualism: Language and Cognition, vol. 10, pp. 51–55. [CrossRef]

de Leeuw, Esther, Monika S. Schmid, and Ineke Mennen. 2010. The effects of contact on native language pronunciation in an L2 migrant context. *Bilingualism: Language and Cognition* 13: 33–40. [CrossRef]

de Leeuw, Esther, Ineke Mennen, and James M. Scobbie. 2012. Singing a different tune in your native language: First language attrition of prosody. *International Journal of Bilingualism* 16: 101–16. [CrossRef]

de Leeuw, Esther, Ineke Mennen, and James M. Scobbie. 2013. Dynamic systems, maturational constraints and L1 phonetic attrition. *International Journal of Bilingualism* 17: 683–700. [CrossRef] de Leeuw, Esther, Aurela Tusha, and Monika S. Schmid. 2018a. Individual phonological attrition in Albanian-English late bilinguals. *Bilingualism: Language and Cognition* 21: 278–95. [CrossRef]

de Leeuw, Esther, Aurela Tusha, Hui Zhao, Kyle Helke, and Alice Greenfield. 2018b. A case of extreme phonetic attrition in the German rhotic. In *Mind Matters in SLA*. Edited by Clare Wright, Thorsten Piske and Martha Young-Scholten. Bristol: Multilingual Matters, pp. 162–83. [CrossRef]

de Leeuw, Esther. 2019. Native speech plasticity in the German-English late bilingual Stefanie Graf: A longitudinal study over four decades. *Journal of Phonetics* 73: 24–39. [CrossRef]

Dmitrieva, Olga. 2019. Transferring perceptual cue-weighting from second language into first language: Cues to voicing in Russian speakers of English. *Journal of Phonetics* 73: 128–43. [CrossRef]

Flege, James E., and Ocke-Schwen Bohn. 2020. The Revised Speech Learning Model. Unpublished Preprint. Available online: https://www.researchgate.net/publication/342923241_The_revised_Speech_Learning_Model (accessed on 20 August 2020).

Flege, James E., Elaine M. Frieda, and Takeshi Nozawa. 1997. Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics* 25: 169–86. [CrossRef]

Flege, James E. 1984. The detection of French accent by American listeners. *Journal of the Acoustical Society of America* 76: 692–707. [CrossRef]

Flege, James E. 1987. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15: 47–65. [CrossRef]

Flege, James E. 1995. Second language speech learning: Theory, findings, and problems. In *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*. Edited by Winifred Strange. Maryland: York Press, pp. 233–77.

Fowler, Carol A., Valery Sramko, David J. Ostry, Sarah A. Rowland, and Pierre Hallé. 2008. Crosslanguage phonetic influences on the speech of French–English bilinguals. *Journal of Phonetics* 36: 649–63. [CrossRef]

Green, David W. 1998. Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition* 1: 67–81. [CrossRef]

Gregg, Kevin R. 2010. Review article: Shallow draughts: Larsen-Freeman and Cameron on complexity. *Second Language Research* 26: 549–60. [CrossRef]

Grosjean, François, and Joanne L. Miller. 1994. Going in and out of languages: An example of bilingual flexibility. *Psychological Science* 5: 201–6. [CrossRef]

Grosjean, François. 1989. Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language* 36: 3–15. [CrossRef]

Grosjean, François. 2001. The bilingual's language modes. In *One Mind, Two Languages: Bilingual Language Processing*. Edited by Janet L. Nicol. Oxford: Blackwell, pp. 1–22.

Guion, Susan G. 2003. The vowel systems of Quichua-Spanish bilinguals: Age of acquisition effects on the mutual influence of the first and second languages. *Phonetica* 60: 98–128. [CrossRef]

Harrington, Jonathan, Sallyanne Palethorpe, and Catherine I. Watson. 2000. Monophthongal vowel changes in Received Pronunciation: An acoustic analysis of the Queen's Christmas broadcasts. *Journal of the International Phonetic Association* 30: 63–78. [CrossRef]

Harrington, Jonathan. 2006. An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts. *Journal of Phonetics* 34: 439–57. [CrossRef]

Hopp, Holger, and Monika S. Schmid. 2013. Perceived foreign accent in first language attrition and second language acquisition: The impact of age of acquisition and bilingualism. *Applied Psycholinguistics* 34: 361–94. [CrossRef]

Illman, Veera, and Päivi Pietilä. 2018. Multilingualism as a resource in the foreign language classroom. *ELT Journal* 72: 237–48. [CrossRef]

Jilka, Matthias. 2009. Assessment of phonetic ability. In *Language Talent and Brain Activity Trends in Applied Linguistics*. Edited by Grzegorz Dogil and Susanne M. Reiterer. Berlin: De Gruyter, pp. 17–66.

Kartushina, Natalia, and Clara D. Martin. 2019. Third-language learning affects bilinguals' production in both their native languages: A longitudinal study of dynamic changes in L1, L2 and L3 vowel production. *Journal of Phonetics* 77: 100920. [CrossRef]

Kartushina, Natalia, Alexis Hervais-Adelman, Ulrich H. Frauenfelder, and Narly Golestani. 2016a. Mutual influences between native and non-native vowels in production: Evidence from short-term visual articulatory feedback training. *Journal of Phonetics* 57: 21–39. [CrossRef]

Kartushina, Natalia, Ulrich H. Frauenfelder, and Narly Golestani. 2016b. How and when does the second language influence the production of native speech sounds: A literature review. *Language Learning* 66: 155–86. [CrossRef]

Kim, Midam, William S. Horton, and Ann R. Bradlow. 2011. Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology* 2: 125–56. [CrossRef]

Kim, Midam. 2009. Phonetic accommodation in conversations between native and nonnative speakers. *Journal of the Acoustical Society of America* 125: 2764. [CrossRef]

Köpke, Barbara, and Monika S. Schmid. 2004. First language attrition: The next phase. In *First Language Attrition: Interdisciplinary Perspectives on Methodological Issues*. Edited by Monika S. Schmid, Barbara Köpke, Merel Keijzer and Lina Weilemar. Amsterdam: John Benjamins, pp. 1–43. [CrossRef]

Krippendorff, Klaus. 2018. *Content Analysis: An Introduction to Its Methodology*, 4th ed. Thousand Oaks: SAGE.

Kupisch, Tanja, Dagmar Barton, Katja Hailer, Ewgenia Klaschik, and Ilse Stangen. 2014. Foreign accent in adult simultaneous bilinguals. *The Heritage Language Journal* 11: 123–50.

Kuznetsova, Alexandra, Per Bruun Brockhoff, and Rune Haubo Bojesen Christensen. 2017. LmerTest: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82: 1–26. [CrossRef]

Lenneberg, Eric H. 1967. *Biological Foundations of Language*. New York: Wiley.

Lewandowski, Natalie, and Matthias Jilka. 2019. Phonetic convergence, language talent, personality and attention. *Frontiers in Communication* 4: 18. [CrossRef]

Littlewood, William, and Baohua Yu. 2011. First language and target language in the foreign language classroom. *Language Teaching* 44: 64–77. [CrossRef]

Lloyd-Smith, Anika, Marieke Einfeldt, and Tanja Kupisch. 2020. Italian-German bilinguals: The effects of heritage language use on accent in early-acquired languages. *International Journal of Bilingualism* 24: 289–304. [CrossRef]

Major, Roy C. 1992. Losing English as a first language. *The Modern Language Journal* 76: 190–208. [CrossRef]

Martínez-Celdrán, Eugenio, Ana Ma Fernández-Planas, and Josefina Carrera-Sabaté. 2003. Illustrations of the IPA: Castilian Spanish. *Journal of the International Phonetic Association* 33: 255–59. [CrossRef]

Mayr, Robert, and Simona Montanari. 2015. Cross-linguistic interaction in trilingual phonological development: The role of the input in the acquisition of the voicing contrast. *Journal of Child Language* 42: 1006–35. [CrossRef] [PubMed]

Mayr, Robert, Sasha Price, and Ineke Mennen. 2012. First language attrition in the speech of Dutch-English bilinguals: The case of monozygotic twin sisters. *Bilingualism: Language and Cognition* 15: 687–700. [CrossRef]

Mayr, Robert, Laura López-Bueno, Martín Vázquez Fernández, and Gisela Tomé Lourido. 2019. The role of early experience and continued language use in bilingual speech production: A study of Galician and Spanish mid vowels by Galician-Spanish bilinguals. *Journal of Phonetics* 72: 1–16. [CrossRef]

Mayr, Robert, Llian Roberts, and Jonathan Morris. 2020. Can you tell by their English if they can speak Welsh? Accent perception in a language contact situation. *International Journal of Bilingualism* 24: 740–66. [CrossRef]

Mennen, Ineke. 2004. Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics* 32: 543–63. [CrossRef]

Montgomery, Chris, and Emma Moore. 2018. Evaluating S(c)illy voices: The effects of salience, stereotypes, and co-present language variables on real-time reactions to regional speech'. *Language* 94: 629–61. [CrossRef]

Moyer, Alene. 1999. Ultimate attainment in L2 phonology: The critical factors of age, motivation, and instruction. *Studies in Second Language Acquisition* 21: 81–108. [CrossRef]

Muldner, Kasia, Leah Hoiting, Leyna Sanger, Lev Blumenfeld, and Ida Toivonen. 2019. The phonetics of code-switched vowels. *International Journal of Bilingualism* 23: 37–52. [CrossRef]

Nodari, Rosalba, Chiara Celata, and Naomi Nagy. 2019. Socio-indexical phonetic features in the heritage language context: Voiceless stop aspiration in the Calabrian community in Toronto. *Journal of Phonetics* 73: 91–112. [CrossRef]

Office for National Statistics. 2017. Living Abroad: Migration between Britain and Spain. Available online: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/articles/livingabroad/migrationbetweenbritainandspain#what-are-spanish-citizens-living-in-the-uk-doing (accessed on 20 August 2020).

Olson, Daniel J. 2013. Bilingual language switching and selection at the phonetic level: Asymmetrical transfer in VOT production. *Journal of Phonetics* 41: 407–420. [CrossRef]

Piccinini, Page, and Amalia Arvaniti. 2015. Voice onset time in Spanish-English spontaneous code-switching. *Journal of Phonetics* 52: 121–37. [CrossRef]

Piske, Thorsten, Ian R. A. MacKay, and James E. Flege. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29: 191–215. [CrossRef]

Qualtrics. 2019. Qualtrics XM Platform™. Computer Program. Available online: http://www.qualtrics.com (accessed on 20 April 2019).

R Core Team. 2018. *R: A Language and Environment for Statistical Computing (Version 3.2.2)*. Computer Program. Vienna: R Foundation for Statistical Computing, Available online: http://www.R-project.org/ (accessed on 20 February 2019).

Sancier, Michele L., and Carol A. Fowler. 1997. Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics* 27: 421–36. [CrossRef]

Sankoff, Gillian, and Hélène Blondeau. 2007. Language change across the lifespan:/r/in Montreal French. *Language* 83: 560–88. [CrossRef]

Schmid, Monika S., and Holger Hopp. 2014. Comparing foreign accent in L1 attrition and L2 acquisition: Range and rater effects. *Language Testing* 31: 367–88. [CrossRef]

Schmid, Monika S., and Barbara Köpke. 2017. The relevance of first language attrition to theories of bilingual development. *Linguistic Approaches to Bilingualism* 7: 637–67. [CrossRef]

Schmid, Monika S. 2007. First language attrition. *WIREs Cognitive Science* 4: 117–23. [CrossRef]

Schmid, Monika S. 2010. Languages at play: The relevance of L1 attrition to the study of bilingualism. *Bilingualism: Language and Cognition* 13: 1–7. [CrossRef]

Schmid, Monika S. 2014. The debate on maturational constraints in bilingual development: A perspective from first language attrition. *Language Acquisition* 21: 386–410. [CrossRef]

Selinker, Larry. 1972. Interlanguage. *International Review of Applied Linguistics in Language Teaching* 10: 209–31. [CrossRef]

Stangen, Ilse, Tanja Kupisch, Ana Lia Proietti Ergün, and Marina Zielke. 2015. Foreign accent in heritage speakers of Turkish in Germany. In *Transfer Effects in Multilingual Language Development*. Edited by Hagen Peukert. Amsterdam: John Benjamins, pp. 87–108. [CrossRef]

Stoehr, Antje, Titia Benders, Janet G van Hell, and Paula Fikkert. 2017. Second language attainment and first language attrition: The case of VOT in immersed Dutch–German late bilinguals. *Second Language Research* 33: 483–518. [CrossRef] [PubMed]

Tobin, Stephen J., Hosung Nam, and Carol A. Fowler. 2017. Phonetic drift in Spanish-English bilinguals: Experiment and a self-organizing model. *Journal of Phonetics* 65: 45–59. [CrossRef] [PubMed]

Turnbull, Miles, and Jennifer Dailey-O'Cain, eds. 2009. *First Language Use in Second and Foreign Language Learning*. Clevedon: Multilingual Matters.

Ulbrich, Christiane, and Ineke Mennen. 2016. When prosody kicks in: The intricate interplay between segments and prosody in perceptions of foreign accent. *International Journal of Bilingualism* 20: 522–49. [CrossRef]

Ulbrich, Christiane, and Mikhail Ordin. 2014. Can L2-English influence L1-German? The case of post-vocalic/r/. *Journal of Phonetics* 45: 26–42. [CrossRef]

languages

MDPI

Article

The Effect of Instructed Second Language Learning on the Acoustic Properties of First Language Speech

Olga Dmitrieva [1,*], Allard Jongman [2] and Joan A. Sereno [2]

1 School of Languages and Cultures, Purdue University, West Lafayette, IN 47907, USA
2 Department of Linguistics, University of Kansas, Lawrence, KS 66045, USA; jongman@ku.edu (A.J.); sereno@ku.edu (J.A.S.)
* Correspondence: odmitrie@purdue.edu

Received: 1 September 2020; Accepted: 20 October 2020; Published: 26 October 2020

Abstract: This paper reports on a comprehensive phonetic study of American classroom learners of Russian, investigating the influence of the second language (L2) on the first language (L1). Russian and English productions of 20 learners were compared to 18 English monolingual controls focusing on the acoustics of word-initial and word-final voicing. The results demonstrate that learners' Russian was acoustically different from their English, with shorter voice onset times (VOTs) in [−voice] stops, longer prevoicing in [+voice] stops, more [−voice] stops with short lag VOTs and more [+voice] stops with prevoicing, indicating a degree of successful L2 pronunciation learning. Crucially, learners also demonstrated an L1 phonetic change compared to monolingual English speakers. Specifically, the VOT of learners' initial English voiceless stops was shortened, indicating assimilation with Russian, while the frequency of prevoicing in learners' English was decreased, indicating dissimilation with Russian. Word-final, the duration of preceding vowels, stop closures, frication, and voicing during consonantal constriction all demonstrated drift towards Russian norms of word-final voicing neutralization. The study confirms that L2-driven phonetic changes in L1 are possible even in L1-immersed classroom language learners, challenging the role of reduced L1 use and highlighting the plasticity of the L1 phonetic system.

Keywords: American English; Russian; voicing; classroom learning; second language acquisition; first language drift

1. Introduction

Cross-linguistic phonetic interaction in bilingualism and language learning is believed to be bidirectional: the earlier acquired, more established language (L1) can be affected by the later acquired, often non-dominant, language (L2). This type of crosslinguistic interaction is known by many names: back-transfer, reverse interference, phonetic drift, and language attrition, to name a few. We define this type of interaction as phonetic changes in speakers' L1 brought about by use of L2 and refer to these changes primarily as L2-to-L1 (phonetic) effects or L1 drift.

A few prominent lines of inquiry dominated the previous research on L2-to-L1 effects, leaving the full scope of this phenomenon under-explored. Specifically, the majority of previous work has focused on proficient bilinguals or advanced second language learners and most speakers were studied in the situation of L2 immersion (Baker and Trofimovich 2005; Barlow et al. 2013; Bergmann et al. 2016; Caramazza et al. 1973; Chang 2012; De Leeuw 2019; De Leeuw et al. 2018; De Leeuw et al. 2010; Flege 1987; Fowler et al. 2008; Guion 2003; Harada 2003; Hopp and Schmid 2013; Kartushina and Martin 2019; Lang and Davidson 2019; Lev-Ari and Peperkamp 2013; MacLeod and Stoel-Gammon 2005; Major 1992; Mayr et al. 2012; Mora and Nadeu 2012; Mora et al. 2015; Sancier and Fowler 1997; Simonet 2010; Tobin et al. 2017; Ulbrich and Ordin 2014). Moreover, language pairings often involved Western European languages, such as English, Spanish, French, German, and Dutch, which tend to be relatively similar phonologically and share the Latin alphabet. Finally, these studies on cross-language interaction have typically focused on sound classes that have distinct phonetic realizations in the respective languages, such as oral stops, distinguished across languages via voice onset time (VOT), or oral vowels, distinguished via formant frequencies.

The current study expands the scope of previous work on L2-to-L1 effects by examining a population of relatively inexperienced learners of a rarely studied Slavic second language (Russian). The L2 learners in the present study reside in the home country and are immersed in their native language (American English). Moreover, in addition to inquiring how comparable phonological categories can affect one another's acoustic realization across languages, the present study also considers the transferability of phonological processes from L2 to L1. We target the acoustic realization of word-initial voiced and voiceless stops in native speech of American learners of Russian, to determine whether it has been affected by exposure to Russian. In addition, we investigate word-final stops, fricatives and affricates in learners' English to establish whether their productions show an effect of the Russian final devoicing rule.

In the following sections, we discuss the theoretical underpinnings of the L2-to-L1 phonetic effects (Section 1.1), provide a brief overview of the previous literature on the topic (Section 1.2), and introduce the details of the present study (Section 1.3).

1.1. Mechanism of L2-to-L1 Effects

Among the theoretical models put forth to account for the production of second language speech, the speech learning model (or SLM/SLM-r; Flege 1995, 2003; Flege and Bohn 2020) explicitly predicts bidirectional phonetic interactions and outlines their general mechanism.

SLM postulates that sound categories of learners' first and second languages coexist in the same phonological space, which a-priori creates a possibility for mutual influence. Moreover, SLM proposes that a mechanism of 'equivalence classification' affects the perception of L2 sounds that are acoustically non-identical but similar to existing L1 categories. As a result, corresponding L1 and L2 sounds are joined under the same category and their acoustic properties are predicted to affect each other, such that L2 sounds are realized in an L1-like manner, and L1 sounds are produced similarly to L2 ones—a situation known as category assimilation.

Flege's own work (e.g., Flege 1987) and much of the subsequent research, however, demonstrated that phonetically separate sound categories are nevertheless maintained across languages in the speech of bilinguals, with one or both deviating from the monolingual norm in the direction of assimilation to the other language (Baker and Trofimovich 2005; Caramazza et al. 1973; Chang 2012; Flege and Eefting 1987a, 1987b; Fowler et al. 2008; Harada 2003; Major 1992; Sancier and Fowler 1997; Sundara and Baum 2006). This cross-language separation suggests that bilinguals are able to discern the acoustic–phonetic differences between the cross-language equivalents even when they are merged under the same category. Moreover, this ability is an important condition of the L2-to-L1 effects. If bilinguals perceived L2 sounds as indistinguishable from the ones in their L1, there could not be any influence of L2 on the production of L1.

This brings us to the importance of sufficient experience with L2 required for L2-to-L1 effects to take place. The assumption that L2 experience plays an important role has dominated most of the literature on L2-to-L1 phonetic effects. A large proportion of previous cross-sectional studies reported that the L1 was affected by the L2 either exclusively or to a greater degree in participants with longer L2 exposure and/or higher L2 proficiency (Baker and Trofimovich 2005; Bergmann et al. 2016; Dmitrieva et al. 2010; Flege 1987; Guion 2003; Herd et al. 2015; Huffman and Schuhmann 2016; Lang and Davidson 2019; Major 1992; Peng 1993; Schmid 2013; Schuhmann and Huffman 2015; Tobin et al. 2017).

A reduction in L1 use may also be responsible for the observed L2-assimilatory changes in the L1 speech of bilinguals (De Leeuw et al. 2010; Kartushina and Martin 2019; Mora and Nadeu 2012; Mora et al. 2015; Sancier and Fowler 1997). Indeed, phonetic changes in L1 were typically detected in circumstances that simultaneously provided greater L2 exposure and limited the use of L1, i.e., in bilinguals who were immersed in the L2-dominant environment. Moreover, in some cases it was proposed that continuous L1 use prevented L1 drift despite intensive L2 exposure (De Leeuw et al. 2010; Tobin et al. 2017).

To summarize, current theoretical models predict the emergence of L2-to-L1 phonetic effects in experienced L2 speakers, with a possible added condition of reduced L1 use. In the following section, we review relevant studies which serve to refine these general predictions.

1.2. Previous Research on L2-to-L1 Phonetic Effects

Research consistently demonstrated that greater L2 experience and proficiency lead to a greater likelihood of L2-to-L1 phonetic effects for sequential bilinguals and adult language learners. For example, Flege (1987) demonstrated that the VOT of English [t] was significantly more French-like in the speech of Americans residing in Paris, compared to American students and teachers of French who were residing domestically. The French [t] of speakers of French residing in Chicago was also significantly different from that of French monolinguals, in the direction of assimilation to English, indicating the combined effect of proficiency, experience, and immersion. Later, Lang and Davidson (2019) showed that only Americans residing in Paris, but not American students on a short-term study abroad in France, experienced a drift in native vowel acoustics in the direction of L2 norms, confirming the important role of long-term immersion.

L2 pronunciation proficiency has also been linked more directly to changes in L1 phonetics. Major (1992) reported a positive correlation between L2 proficiency and L1 drift for American immigrants to Brazil: the closer they approximated Portuguese VOT norms in their production of L2 voiceless stops, the more they deviated from native norms in their L1 productions, in the direction of assimilation to L2 (although see Kartushina and Martin (2019) who report a negative L2 proficiency–L1 drift correlation).

The age of L2 acquisition also plays a role in promoting L2-to-L1 phonetic effects. In Guion (2003), only early and mid but not late Quechua–Spanish bilinguals revealed an effect of L2 (Spanish) on native vowel acoustics. Similarly, in Baker and Trofimovich (2005) an L2-to-L1 cross-language influence in vowels was uncovered only for early but not late Korean–English bilinguals, suggesting the importance of accumulated L2 experience.

Estimating L2 experience via length of residence in an L2-dominant environment, Dmitrieva et al. (2010) showed that Russian speakers of English with greater L2 experience were more likely to realize final obstruents in Russian in a more English-like manner, with less devoicing. Similarly, Bergmann et al. (2016) established that native speech of long-term German immigrants to Canada was perceived as more accented by their monolingual compatriots as a function of a longer residence abroad.

While the effects of L2 exposure and L2 proficiency are almost inevitably conflated in research on long-term immigrants, Chang (2012, 2013) was able to disentangle the two. Chang (2012) demonstrated L1 drift in several phonetic parameters, including VOT and vowel spectrum, in beginner American learners of Korean after only a short immersion in Korean during a study abroad program. Crucially, these participants achieved only elementary proficiency in Korean by the end of the six-week program, while L1 drift was observed already at week two. This work suggests that L2 proficiency by itself is not a necessary condition of L2-to-L1 phonetic effects, but L2 exposure which comes about due to L2 immersion may be.

Chang's work does share the element of L2 immersion, providing a level of L2 input that is both abundant and authentic, even if it is primarily overheard, with much of the previous literature. There is evidence that even overhearing-type exposure to another language may have important and long-lasting consequences. Au et al. (2002) and Knightly et al. (2003) showed that individuals who, early in life, were exposed to Spanish without learning it ('overhearers'), upon enrolling in an L2 Spanish class, demonstrated near-native like VOTs in Spanish, compared to learners in the same class who did not overhear Spanish earlier in life. Moreover, Chang (2019b) demonstrated that L1 drift persisted for L2 learners immersed in L2 even when they no longer actively used the second language, suggesting that ambient language exposure in adulthood as well as in early childhood is an important factor affecting language production. Caramazza et al. (1973) also reported an interaction between English and French, exclusively in the direction of English affecting French, for residents of Canada who spoke only one of the two languages but were presumably exposed to both.

This work raises an important question about the minimum amount of L2 exposure required to trigger L2-to-L1 phonetic effects. Clearly, the intensive and authentic exposure provided by L2 immersion can be sufficient. However, what about non-immersion-type exposure to L2? Research on L2 learners in non-immersion situations is relatively scarce but it provides some indication that an even more fleeting introduction to another language may trigger phonetic changes in L1.

Traditional classroom learners of additional languages have been largely overlooked when it comes to L2-to-L1 effects. A number of studies examining non-immersion population were often conducted with small numbers of participants, thus arriving at somewhat inconclusive results. Huffman and Schuhmann (2016) examined four beginner American learners of Spanish and reported little evidence of L2-to-L1 phonetic effects. Between weeks 2 and 6 of language instruction, learners demonstrated no changes in the VOT of native voiced or voiceless stops. Only the frequency of prevoicing in English suggested a tendency to dissimilate away from Spanish: three participants decreased or eliminated prevoicing from their English voiced stops. Schuhmann and Huffman (2015) did show that after a period of explicit phonetic training, three out of five learners of Spanish shortened their English voiceless stops' VOT, indicating assimilation to Spanish. Herd et al. (2015), the only large-scale (N = 40) cross-sectional study of classroom learners known to us, demonstrated that near-native and advanced learners of Spanish produced English voiced stops with more negative VOTs than beginner learners. The near-native, advanced, and intermediate learners also produced more peripheral English vowels than beginner learners did—a difference also compatible with the effect of Spanish on English. This study indicates that L1 drift is possible in more experienced classroom L2 learners but, in the absence of the monolingual control group, it was not established whether beginner learners also modified the acoustics of their native speech in the direction of second language norms. Overall, the available studies on L2-to-L1 effects in classroom learners provide limited evidence that L2 exposure in the classroom may be sufficient to trigger L1 drift.

Another reason to study classroom learners is the fact that they continue to reside in the home country while acquiring their L2. Most foreign language courses in US colleges provide active instruction 3–5 h a week. For the remainder of the time, learners use their L1. The amount of reduction in L1 use and exposure is most likely negligible in these circumstances.

This aspect of classroom language learning is important because the reduction in L1 use associated with L2 immersion could play an important role in creating conditions for the L2-to-L1 phonetic effects. Conversely, continued L1 use has been suggested to promote and protect the 'authenticity' of L1 speech (Kartushina et al. 2016b). For example, Bergmann et al. (2016) demonstrated a negative correlation between the amount of L1 use and the degree of perceived non-native accent in the L1 speech of long-term German immigrants to North America. De Leeuw and colleagues also showed that the German of immigrants to Canada and The Netherlands was less likely to be perceived as non-native sounding if they had a high amount of contact with other Germans in a monolingual mode (De Leeuw et al. 2010). Moreover, Mora and colleagues (Mora and Nadeu 2012; Mora et al. 2015) reported that greater use of L1 Catalan promoted more monolingual-like Catalan vowels in Catalan–Spanish bilinguals. Although Tobin et al. (2017) did not detect any L1 drift in the native speech of Spanish learners of English after a 3–4 months period of L2 immersion in the United States, they explained this result by the lack of a sufficient reduction in L1 use.

The dominant, and thus more frequently used, language is also believed to be protected from the cross-linguistic influence. For example, Kartushina and Martin (2019) showed that, in balanced Catalan–Spanish bilinguals, both languages were affected by immersion in English but in Spanish-dominant bilinguals only Catalan vowels drifted towards English (see also Caramazza et al. (1973) and Mack (1989)).

To summarize, much previous research indicates that while advanced L2 proficiency is not a necessary condition for L2-to-L1 phonetic effects, greater L2 exposure and experience promote L1 drift. Immersion-type exposure to L2 is particularly conducive to L1 drift. Moreover, the reduction in L1 use, which typically co-occurs with L2 immersion and L2 dominance, is another possible condition for L2-to-L1 phonetic effects.

The population of classroom learners, which has not been widely studied with respect to L1 drift, provides an essential complement to previous work on immersed learners; a comparison that leads to a better understanding of the role of L2 immersion and reduced L1 use in bidirectional cross-language interaction. The following section describes the present study designed to address the question of L1 drift in classroom language learners.

1.3. Present Study

The present study aims to determine whether exposure to a second language via classroom learning can lead to phonetic changes in the native speech of the learners. The second language studied by our participants is Russian.

Russian is a relatively unusual choice for American learners and a comparatively difficult language to acquire for English speakers. In a ranking of languages encompassing four different difficulty categories, the US Foreign Service Institute placed Russian in category III, among 'hard' languages with significant linguistic and/or cultural differences from English (https://www.state.gov/foreign-language-training/), and specified that approximately 1100 class hours are required to reach general professional proficiency in speaking and reading (S3 and R3). This amounts to 14 semesters of study, assuming a fairly typical five hours per week over a 16-week semester study pattern. Thus, although participants for the present study were recruited from the second through to the sixth semesters of Russian study, it is reasonable to assume that most had not managed to reach advanced proficiency in this amount of time.

Unlike more frequently studied languages such as French, German, Italian, and Spanish, Russian does not share the same writing system with English. This makes L1 English–L2 Russian a qualitatively different and novel language pairing to consider. In particular, we ask whether, L1 drift is as likely in pairings of languages with fewer linguistic, orthographic, and cultural similarities as among more similar languages.

We consider the voice onset time of word-initial voiced and voiceless stops as the phonetic aspect potentially subject to L1 drift. In addition to this commonly studied parameter, we examine onset f0—pitch at the beginning of the post-consonantal vowel—as a secondary correlate of voicing. Secondary correlates have rarely been studied in L2 learners and we know little about their propensity to drift towards L2 in L1 speech.

Russian realizes its initial prevocalic [+voice] stops as robustly prevoiced (with negative VOT) and its initial prevocalic [−voice] stops as voiceless unaspirated (short lag VOT) (Ringen and Kulikov 2012). English realizes its initial prevocalic [+voice] stops as a combination of weakly prevoiced (about 30% for the population, Dmitrieva et al. 2015) and voiceless unaspirated stops (70%), and its initial prevocalic [−voice] stops as voiceless aspirated (long lag VOT) (Lisker and Abramson 1964). This phonetic difference between Russian and English stop voicing is usually not taught explicitly in Russian language courses, as was confirmed by Purdue University Russian language instructors.

The expected pattern of L1 drift, based on previous research, includes a well-documented tendency towards VOT shortening in voiceless English stops. It is also possible that the prevoicing period in English [+voice] stops could be lengthened under the influence of Russian. Finally, the proportion of prevoiced to voiceless unaspirated stops among English [+voice] segments could change towards a greater frequency of prevoicing, in assimilation with Russian.

With respect to onset f0, the two languages demonstrate a congruent covariation of f0 with phonological categories (lower f0 after [+voice] stops) but an incongruent covariation with phonetic VOT categories: first, f0 is lower after prevoiced stops than after voiceless unaspirated stops in Russian but there is no such difference in English because short lag and lead VOT stops are variants of the same phonological category (Kulikov 2012; Dmitrieva et al. 2015). Thus, exposure to Russian could lead to f0 lowering after prevoiced stops in participants' English speech. Second, English voiceless unaspirated stops are characterized by low onset f0, as they are phonologically voiced, while Russian voiceless unaspirated stops are characterized by high onset f0, as they are phonologically voiceless. Thus, an L2-to-L1 effect in this case would involve the relative raising of onset f0 after voiceless unaspirated stops in the English of Russian learners.

Finally, we investigate the temporal indices of voicing in word-final obstruents: preceding vowel duration, consonant constriction duration, and duration of voicing during constriction. This additional area of interest was selected because of important differences between English and Russian in the way phonological and phonetic voicing is treated in final obstruents. English, for the most part, maintains phonetic differences between phonologically voiced and voiceless final obstruents, although there is a gradient tendency to devoice in this position, especially for fricatives (Davidson 2016). Russian, on the other hand, features categorical devoicing in word-final position. We aim to investigate the possibility of L2-to-L1 influence on the basis of phonological rules which apply in the L2. We hypothesize that learners' L1 may adopt this phonological process from the L2 (Barlow et al. 2013; Simonet and Amengual 2020).

We further hypothesize that such influence may be especially likely for areas of L1 phonology that trend towards change, in particular if change is in the direction of the L2 process, in this case, devoicing (see Barlow et al. (2013) and Bullock and Gerfen (2004) for similar reasoning). Thus, English speakers exposed to Russian may be expected to demonstrate a stronger tendency to devoice in word-final position than is observed for monolingual English speakers.

To summarize, the present study examines L1-immersed classroom language learners in order to extend previous investigations of L2-to-L1 effects to populations not characterized by extensive L2 exposure and reduced L1 use due to L2 immersion. To establish the phonetic effects of their L2, Russian, on their L1, English, we examine the acoustic properties of word-initial stops (VOT and onset f0), and word-final obstruents (temporal indices of final voicing).

Following previous research, we conduct two types of comparisons: that between learners' L1 and L2, in order to determine whether the two systems are distinct or merged with respect to the select acoustic properties (a within-subject comparison) and those between learners' and monolinguals' L1s, in order to determine whether L2-to-L1 effects have taken place in learners' speech (a between-subject comparison). We believe that it is important to conduct both comparisons in order to demonstrate that a degree of phonetic learning has taken place in these speakers' L2, and that L1 drift, if present in their speech, is consistent with the nature of phonetic learning they achieved in their L2 speech. By establishing the degree of L2 phonetic learning for our participants, we further our understanding of the conditions under which L1 drift can be expected to occur. Moreover, the cooccurrence of L1 drift and L2 phonetic learning for the same features supports the notion that L2 phonetic learning is what triggers L1 drift.

To determine that L1 drift is a relatively stable feature of learners' native speech as opposed to the short-term effect of producing speech in the two languages in immediate succession, we analyzed the effect of the order of language elicitation.

We also examine the relationship between the extent of individual L1 drift and L2 proficiency in order to test the hypothesis that magnitude of drift in L1 is linked to the degree of pronunciation gains in L2.

Thus, the three main objectives of the present research are: (1) to determine whether phonetic learning has taken place in the Russian speech of learners; (2) to determine whether L1 drift has taken place in the English speech of learners; and (3) to determine whether the degree of phonetic learning/pronunciation gains were correlated with the degree of L1 drift.

2. Materials and Methods

2.1. Participants

Twenty native speakers of American English learning Russian as a second language participated in the study: eleven men and nine women, between the ages of 19 to 24 years (M = 20.6, SD = 1.3). They were recruited and recorded in two locations: Purdue University (14 participants) and the University of Kansas (6 participants). Participants filled out a language background questionnaire after the recording. All reported English as their first and native language. All participants reported learning Russian mainly through college classroom instructions and only four participated in a 2–4 months-long Russian study abroad program some time during the year preceding their enrollment in the study. On average, they studied Russian for 5 semesters by the time of participation (SD = 3, R = 1.5–12). The amount of class time varied by level, e.g., from five hours a week for semesters 1 through 4 of Russian, to three hours a week, starting from the 5th semester (Purdue campus).

Participants reported using Russian mostly in class or with classmates, on average for four hours per week (ranging from one to 6 h). Four participants reported using Russian with a family member but only up to one hour a week. The most commonly reported type of engagement with Russian was reading (M = 2 h/week, R = 1–6 h/week). Writing in Russian was the second most common activity (M = 2 h/week, R = 0.5–4 h/week). Only about half of the participants reported listening to Russian radio or watching Russian TV (M = 3 h/week, R = 1–6 h/week).

Participants' average self-reported Russian fluency was 'fair' ('3' on a 7-point scale), and the degree of accentedness in Russian was 'moderate' ('3' on a 7-point scale). All participants studied additional modern languages in classroom settings (the majority of participants studied only one additional language per person), most commonly Spanish, French and German (for 5 semesters on average, across these three languages). Achieved proficiency was 'fair' on average ('3' on a 7-point scale). Only three participants reported 'good' or 'very good' knowledge of an additional language (German and Spanish).

Eighteen native speakers of American English from the same dialectal area (Midwest) participated in the study as the control group: four men and fourteen women, between the ages of 18 and 57 (M = 25.8, SD = 9.8). These participants were recruited at Purdue University from the same undergraduate student population. They self-identified as native and monolingual speakers of Midwestern English without significant knowledge of other languages. Although all had some experience of learning a second language in instructional settings (most often Spanish or French), this experience was current or recent for only three participants.

None of the participants in either experimental or control group reported a hearing or speech impairment, and all were compensated for participation with course credit or cash. The study was approved by the Purdue University and University of Kansas Institutional Review Boards, protocols 1409015219 and 00003743, respectively.

2.2. Elicitation Materials

Elicitation materials consisted of English and Russian minimal and near-minimal monosyllabic pairs contrasting word-initial and word-final voicing.

The 44 English pairs consisted of 18 stop-initial (e.g., cap–gap), 18 stop-final (e.g., mop–mob), 6 fricative-final (e.g., safe–save), and 2 affricate-final (e.g., rich–ridge) pairs. There was a total of 75 experimental items (some words were used in the word-initial and the word-final condition). Bilabial, alveolar, and velar stops were represented in equal numbers and final fricatives were labiodental (2 pairs) and alveolar (4 pairs). Preceding and following context was largely limited to the vowels [æ], [ɑ], and [ʌ]. There was no significant difference in lexical frequency between voiced and voiceless members of the pairs (COCA Corpus, Davis (2008)). Forty-eight mono- and disyllabic distractor items were also included. A complete list of English target stimuli is provided in Appendix A, Tables A1 and A2.

The 42 Russian pairs consisted of 18 stop-initial (e.g., [kostj]–[gostj] 'bone'–'guest'), 18 stop-final (e.g., [xrjip]–[grjib̥] 'wheeze'–'mushroom'), and 6 fricative-final pairs (e.g., [rjis]–[prjiz̥] 'rice'–'prize'), for a total of 84 experimental items. Bilabial, dental, and velar stops were represented in equal numbers and final fricatives were labiodental (1 pair) or alveolar (4 pairs). Preceding and following vowels were mid-low [e], [a], and [o] in about two-thirds of cases, the rest contained high vowels [i], [u], or [ɨ]. There were no significant differences in lexical frequency between voiced and voiceless stimuli (Russian National Corpus 2003). Forty-five mono- and disyllabic distractor items were also included. A complete list of target stimuli is provided in Appendix A, Tables A3 and A4.

2.3. Procedure

Participants recorded at Purdue University were seated in front of the computer screen in a double-walled sound-attenuated booth. E-prime 2.0 (Psychology Software Tools, Pittsburgh, PA) was used to display the words for elicitation. The words appeared on the screen one by one, in a random order. Each word stayed on the screen for 2 s and was followed by 0.5 s of blank screen. Participants were instructed to pronounce each word the way they speak normally. The whole list was presented three times to each participant with short breaks offered between the blocks. The recording was performed using an Audio-Technica AE4100 cardioid microphone and a TubeMP preamp connected directly to a PC.

For participants recorded at the University of Kansas, a similar procedure was used. PowerPoint software was used to present the prompts on the screen, in a random order for each participant, with each word displayed on the screen for 1.5 s, followed by 1.5 s of blank screen. Recordings were performed in an anechoic chamber, using an Electro-Voice N/D 767a microphone and Marantz PMD671 digital recorder.

This computer-controlled stimulus presentation elicits an appropriately consistent rate of speech across and within participants. The order of Russian and English conditions was counterbalanced across participants, with a brief break between conditions. Due to technical issues, only one repetition of each item was recorded for one experimental participant, and only English data were collected from another experimental participant.

2.4. Measurements

For initial stops, voice onset time (VOT) and onset f0 were measured. For final obstruents, preceding vowel duration, duration of consonantal constriction, and duration of voicing during constriction were measured. Segmentation was performed manually based on Praat (Boersma and Weenink 2018) waveform, and spectrogram representations and using standard segmentation criteria. Measurements were collected using custom-written Praat scripts.

VOT was measured from the onset of consonantal release until the onset of voicing. Onset f0 was measured at the vowel onset as soon as the Praat autocorrelation algorithm detected periodicity. Obtained f0 values were examined for algorithm errors and corrected manually if necessary. Normalization was performed by converting f0 values to semitones relative to each participant's individual mean onset f0, using the formula 12ln(x/individual mean onset f0)/ln2, based on the semitone normalization procedure in Boersma and Weenink (2018). After normalization, outliers more than two standard deviations away from the normalized grand mean onset f0 were removed from further analysis (97% of onset f0 measurements were retained). The resulting values represented the deviation of each onset f0 value, on the logarithmic scale, from each participant's mean, now represented as 0.

Duration of the preceding vowel, duration of the closure for stops/affricates, frication portion for fricatives/affricates, and duration of voicing during constriction were measured for final obstruents.

3. Results

All the reported Linear Mixed Models (LMM) were implemented in SPSS 26.0 with the same random effects structure: a random intercept for subject and for item. Significance of the fixed factors and interactions was assessed via ANOVA tests. All pairwise comparisons were performed with Sidak correction. To avoid averaging across positive and negative VOT values, separate statistical models were fit to stops with prevoicing and stops with positive VOT.

We report results for initial stops first, followed by results for final obstruents. Within each of those sections, we begin by reporting the comparison between learners' Russian and English speech, to determine whether the two languages were produced by learners in a phonetically distinct way and to establish the degree of phonetic learning in their L2. We then proceed to report the comparison between learners' English speech and English speech of monolingual controls to test for L1 drift in learners' speech. We finish by reporting the correlations between the degree of phonetic learning in each learner's L2 and the magnitude of L1 drift in his or her English speech.

3.1. Initial Stops

3.1.1. Learners' Russian vs. English

The goal of this analysis was to establish whether learners' Russian productions were acoustically distinct from their own English speech.

Positive VOT: Positive VOT of initial stops was analyzed using an LMM with Language (Russian vs. English), Voicing, Place of Articulation (included to account for systematic variability in VOT duration as function of place of articulation) and the two-way Language by Voicing interaction as fixed factors. The results demonstrated a significant effect of Language, $F(1, 66.85) = 12.45$, $p = 0.001$, Voicing, $F(1, 66.65) = 690.39$, $p < 0.001$, Place of Articulation, $F(2, 66.46) = 34.13$, $p < 0.001$, and a significant Language by Voicing interaction, $F(1, 66.58) = 24.70$, $p < 0.001$.

The effects of Voicing and Place of Articulation were due to a longer positive VOT for voiceless than for voiced stops and an increase in VOT in the following order: labial < coronal < dorsal, where every pairwise comparison was statistically significant. The effect of Language was due to longer VOTs in English (M = 45 ms, SD = 30 ms) than in Russia (M = 43 ms, SD = 28 ms). This effect was mostly driven by differences between the voiceless stops, while voiced stops were produced with more comparable VOTs across languages (see Figure 1). The significant Language by Voicing interaction confirms the magnitude of this asymmetry. The shortened VOT of initial voiceless stops produced in Russian indicated that learners were in the process of acquiring the phonetics of the Russian voiceless category, by targeting shorter lag productions. However, with a mean of 57 ms their Russian realizations were only 13 ms away from their English long lags (M = 70 ms) and still far from being true Russian-like voiceless stops.

Given that Russian voiceless productions were, on average, shorter in VOT than English ones, another important question is how many instances of learners' Russian stops could be categorized as 'short lag'. Using a relatively generous cut-off of 40 ms (to accommodate for the lower rate of speech in isolated word production and the longer VOT of velar stops) to demarcate the boundary between short lag and long lag voiceless stops (Lisker and Abramson 1964), we calculated the proportions of such realizations in participants' Russian and English speech. Figure 2 shows the distribution. While only 5% of short lags were detected among English voiceless stops, in Russian the proportion rose to 28%, indicating that appreciable VOT shortening affected almost a third of Russian productions. This asymmetry was significant in a chi-square test, $\chi 2(1, N = 2155) = 200.61$, $p < 0.001$.

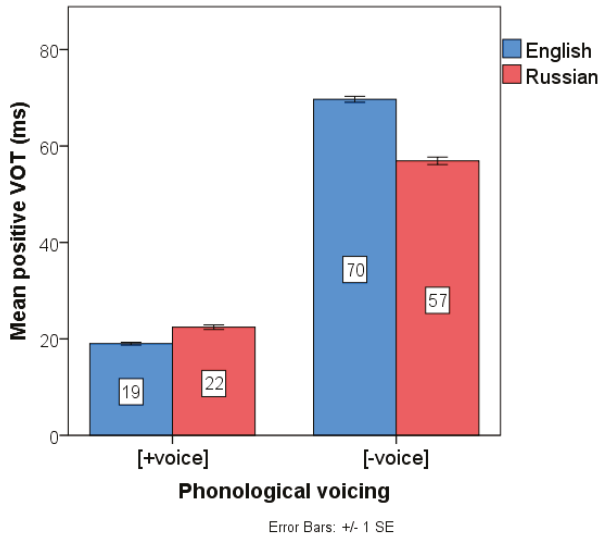


Figure 1. Mean positive voice onset time (VOT) of voiced and voiceless initial stops in learners' English and Russian.

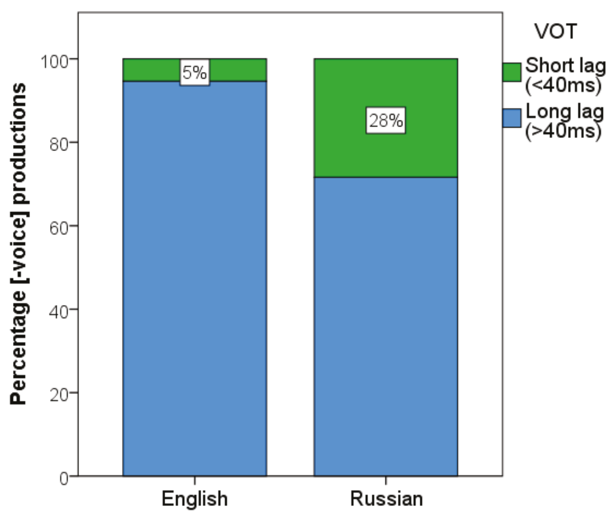


Figure 2. Percentage of short lag productions among [−voice] stops in learners' English and Russian.

Negative VOT: Negative VOT of initial [+voice] stops was analyzed using an LMM with Language and Place of Articulation as fixed factors to establish whether the duration of prevoicing differed between learners' Russian and English speech. The results showed a significant effect of Language, $F(1, 101.85) = 21.10$, $p < 0.001$. Russian [+voice] stops were characterized by a longer prevoicing period (M = 100 ms, SD = 37 ms) than English initial [+voice] stops (M = 78 ms, SD = 27 ms). Thus, although both English and Russian license prevoiced stops as representatives of the [+voice] category, they were phonetically distinct in the realizations of these American learners of Russian.

The frequency of prevoicing is another cross-linguistically distinguishing aspect, since all Russian [+voice] stops are supposed to be produced with prevoicing. To determine the extent to which learners reached this objective, we calculated the proportion of prevoiced realizations among [+voice] Russian and English stops, shown in Figure 3. While only 6% of English voiced stops were realized with prevoicing, 33% of Russian realizations were prevoiced, suggesting that learners were producing prevoiced realizations of Russian [+voice] stops, although well below the rate of native speakers.

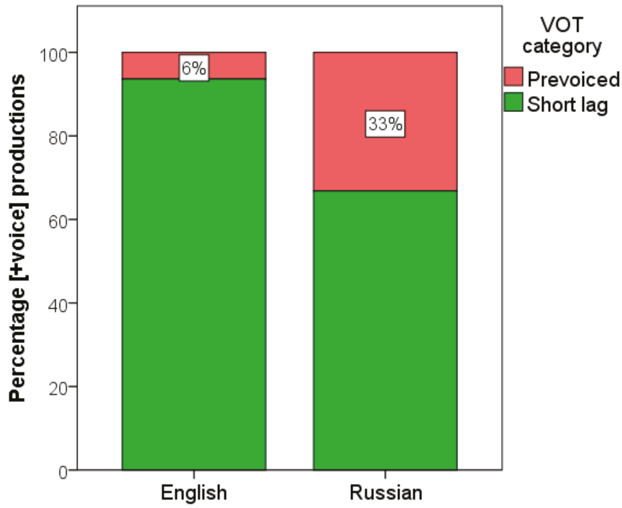


Figure 3. Percentage of prevoiced productions among [+voice] stops in learners' English and Russian.

Onset f0: Figure 4 demonstrates the distribution of normalized onset f0 values in the English and Russian speech of the same participants. A few differences between the languages are apparent. First, prevoiced stops form a more substantial category in Russian than in English, with the distribution visibly shifted towards lower f0 values, compared to the English prevoiced distribution. Second, the Russian [−voice] distribution is less compact than the English one in terms of VOT range, encompassing a span of shorter values (up to 0 ms VOT) and, as a result, overlapping with [+voice] stops produced at short lags. The two distributions (Russian [+voice] short lags and Russian [−voice]) nevertheless maintain a separation in terms of f0 values, with visibly lower f0 of [+voice] short lags.

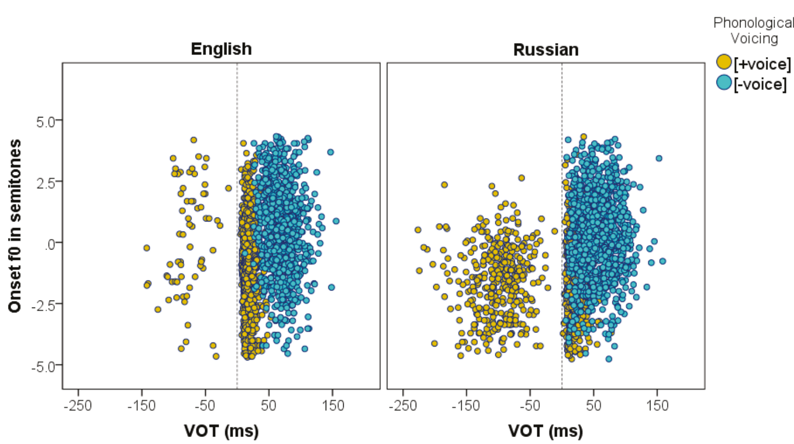


Figure 4. Scatterplots of onset pitch at the beginning of the post-consonantal vowel (f0) and VOT values for learners' English and Russian initial stops.

To examine the alignment of onset f0 values with both the phonological voicing and VOT categories in the two languages of learners, their normalized f0 values were analyzed in an LMM with Language, Voicing, and Language by Voicing interaction as fixed factors. In this analysis, Voicing was a hybrid category with three levels ([+voice] stops were split into those with prevoicing and those without): [+voice] prevoiced, [+voice] short lag, and [−voice]. This was motivated by the fact that the two types of [+voice] stops have very distinct VOT implementations and may be expected to behave differently in Russian where prevoiced stops form a separate phonological category to the exclusion of short lag stops.

The results demonstrated a significant effect of Language, $F(1, 104.85) = 25.52$, $p < 0.001$, and Voicing, $F(2, 192.07) = 104.79$, $p < 0.001$, and a significant Language by Voicing interaction, $F(2, 190.10) = 4.48$, $p = 0.013$. Onset f0 was significantly lower in Russian than in English. The effect of Voicing was driven by significantly higher onset f0 after voiceless stops compared to either prevoiced or short lag [+voice] stops, without a significant difference between the latter. The interaction between Language and Voicing was triggered by the divergent behavior of f0 after prevoiced stops. As shown in Figure 5, Russian prevoiced stops lowered f0 even more than English prevoiced stops.

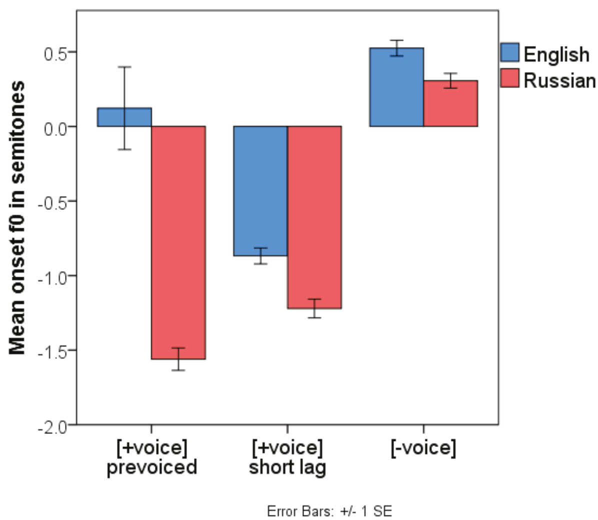


Figure 5. Mean onset f0 across the three categories of initial stops in learners' English and Russian.

To investigate this tendency further, we compared prevoiced and [+voice] short lag stops in English and in Russian separately, in LMM analyses with a single fixed effect: VOT category. The difference was significant only in Russian, $F(1, 1068.69) = 25.63$, $p < 0.001$, where prevoiced stops triggered lower onset f0 than short lag stops, being members of the same phonological category. This result suggests that learners of Russian were developing an awareness of prevoicing as a separate phonological category in Russian and attempting to single it out with a distinct f0 pattern.

We were further interested in onset f0 of [−voice] short lags. The question of interest here is whether, when producing Russian short lags, learners transfer all the co-varying properties of English initial short lags, including low f0.

We compared f0 values of Russian [+voice] short lags, Russian [−voice] short lags (<40 ms), and Russian [−voice] long lags (>40 ms) in an LMM model with a single fixed factor with these three levels. The effect was significant, $F(2, 70.33) = 34.83$, $p < 0.001$, and the results of pairwise comparisons demonstrated that f0 was significantly higher after both voiceless categories than after the voiced one, without a significant difference between the two voiceless categories.

Interim summary: The results demonstrated that learners were attempting to approximate Russian phonetic norms by producing (a) shorter VOTs in Russian [−voice] stops, (b) longer prevoicing in Russian [+voice] stops, (c) more instances of [−voice] stops with short lag VOT in Russian than in English and (d) more instances of [+voice] stops with prevoicing in Russian than in English. The acoustics of learners' Russian stops were significantly different from their English stops. However, they were clearly not reaching native-like phonetic norms (all short lag [−voice] stops and all prevoiced [+voice] stops).

Onset f0 findings indicate that learners were able to manipulate the two correlates of voicing—VOT and onset f0—separately from each other. In particular, they did not transfer the low onset f0 associated with initial short lag stops in English to Russian when producing Russian short lags. Instead, they assigned onset f0 values in accordance with the phonological membership of the intended stop, equally successfully in Russian and in English. One result that deserves special notice is the significantly lower onset f0 assigned to Russian prevoiced [+voice] stops compared to Russian [+voice] short lags. These two sets of realizations were not distinguished via onset f0 in native English; thus, the difference is specific to the Russian productions of learners.

3.1.2. Learners' English vs. Monolingual Controls' English

The goal of this analysis is to determine whether learners' productions of initial English stops were affected by exposure to Russian. This effect would be revealed if significant differences were demonstrated in the acoustic realization (in terms of VOT and onset f0) of initial English stops by the two speaker groups (learners' English vs. monolingual English).

Positive VOT: Positive VOT of initial stops was analyzed using an LMM with Group (Learners vs. Monolinguals), Voicing, Place of Articulation, and Group by Voicing interaction as fixed factors. The results demonstrated a significant effect of Group, $F(1, 35.87) = 6.73$, $p = 0.014$, Voicing, $F(1, 32.64) = 2364.47$, $p < 0.001$, PA, $F(2, 32.43) = 33.75$, $p < 0.001$, and a significant Group by Voicing interaction, $F(1, 3828.06) = 182.54$, $p < 0.001$. The effects of Voicing and Place of Articulation demonstrated a longer VOT for voiceless stops and an increase in VOT in the following order: labial < coronal < dorsal, where every pairwise comparison was statistically significant.

The effect of Group demonstrated a longer overall VOT produced by monolingual participants (M = 59 ms, SD = 35 ms) than by learners (M = 45 ms, SD = 30 ms) (this effect was driven primarily by voiceless stops). The significant interaction between Group and Voicing was due to the fact that voiced stops were produced with comparable VOT values across the two groups, while voiceless stops had longer VOT for monolinguals than for learners. Moreover, as Figure 6 shows, learners' mean voiceless VOT in English was situated between that of monolinguals and their own Russian productions. The shortened voiceless VOT of learners' English is compatible with an influence of Russian, where the voiceless category is realized via short lag VOT.

To assess the possible role of elicitation order, an LMM was conducted on English data from learners only, with Order of language elicitation (Russian first or English first), Voicing, and Voicing by Order as fixed effects (all subsequent analyses of Order were conducted with the same model structure). The results confirmed the effect of Voicing, $F(1, 34.075) = 658.189$, $p < 0.001$, but showed no main effect of Order. The Voicing by Order interaction was significant, $F(1, 2027.94) = 15.89$, $p < 0.001$, due to the fact that in the Russian-first condition, learners' [+voice] English stops were pronounced with longer VOT than in the English-first condition. This result agrees with the observation that learners produced relatively long VOTs for short lag [+voice] stops in Russian (see Figure 6), thus their Russian pronunciation tendencies for these types of stops spilled over into English when it was spoken next. The results therefore revealed no evidence that drift towards shorter VOTs in learners' English was triggered by speaking Russian immediately prior to speaking English.

We were also interested in assessing how many of the English voiceless stops produced by learners could be categorized as 'short lag' as a result of this drift. We used a cut-off value of 40 ms, which categorized 99% of the voiceless stops produced by control speakers as long lags. Interestingly, as shown in Figure 7, the proportion of short lags was slightly higher for learners than for monolinguals. This asymmetry was significant in a chi-square test, $\chi2(1, N = 2028) = 30.24$, $p < 0.001$. Thus, about 5% of stops produced by learners were on the margins of the long lag category, moving into the short lag territory.

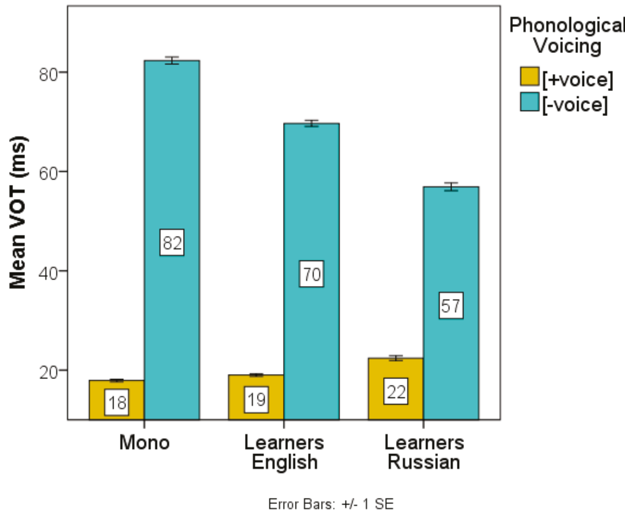


Figure 6. Mean positive VOT of voiced and voiceless initial stops in the English of monolingual controls and learners; learners' Russian is provided for comparison.

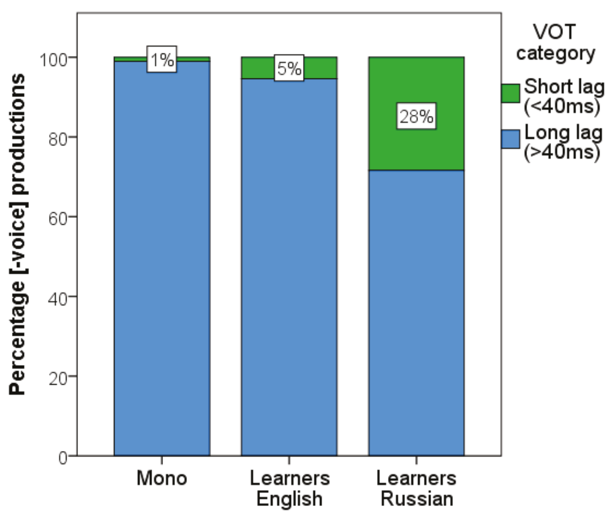


Figure 7. Percentage of short lag productions among [−voice] stops in monolingual controls' and learners' English; learners' Russian is provided for comparison.

Negative VOT: Negative VOT of initial stops was analyzed using an LMM with Group and Place of Articulation as fixed factors to determine whether the duration of prevoicing in learners of Russian differed from that of monolingual controls. Neither factor was a significant predictor of prevoicing duration.

VOT categories in voiced stops: Figure 8 shows that monolingual controls produced prevoicing with almost equal frequency as learners did in their Russian productions (about 30%). In comparison, learners' English was almost devoid of prevoicing, with only 6% prevoiced stops. This asymmetry between controls and learners' English was significant in a chi-square test: $\chi 2(1, N = 2230) = 158.38$, $p < 0.001$.

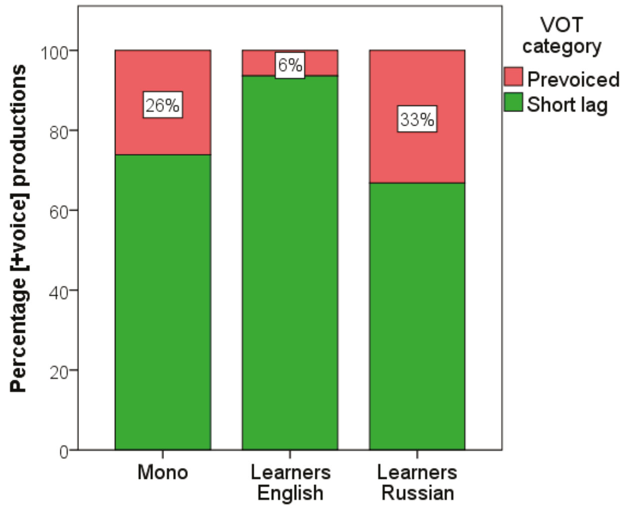


Figure 8. Percentage of prevoiced productions among [+voice] stops in monolingual controls' and learners' English; learners' Russian is provided for comparison.

Moreover, when learners' English productions were split by order of language elicitation, the Russian-first condition resulted in only 3% prevoicing in English, compared to 9% in the English-first condition. These results point towards a possibility of divergence from Russian in learners' English speech; a divergence which can be amplified if Russian is elicited first. Exposure to Russian, where prevoicing marks a separate phonological category, led learners to decrease the incidence of allophonic prevoicing in their English speech.

Onset f0: Figure 9 demonstrates the distribution of onset f0 values and VOT values in English speech of learners and monolingual controls. The two groups present relatively comparable pictures, with the exception of a more substantial prevoiced distribution in controls than in learners and a greater separation between positive VOT categories in monolinguals than in learners. As discussed above, these tendencies are likely due to learners' drift towards shorter VOT in their voiceless productions due to convergence with Russian, and a decrease in the incidence of prevoicing as an expression of divergence from Russian. Learners also demonstrate greater variability in onset f0 values, especially in [−voice] long lag stops.

Onset f0 values in English productions of learners and monolingual controls were compared in an LMM analysis with Group and Voicing (three categories: [+voice] prevoiced, [+voice] short lag, and [−voice]) and Group by Voicing interaction as fixed factors. The results showed a significant effect of Voicing, $F(2, 108.63) = 108.10$, $p < 0.001$, due to a significant difference between [−voice] stops and the two [+voice] categories, with no significant difference between the latter two. No other effects or interactions reached significance. This result suggests that experience with Russian did not significantly affect the way learners realized onset f0 in their English speech.

Interim summary: The results revealed significant differences between learners and monolinguals which can be attributed to convergence and divergence with Russian in learners' English. These differences affected only VOT, while English onset f0 was not affected by exposure to Russian.

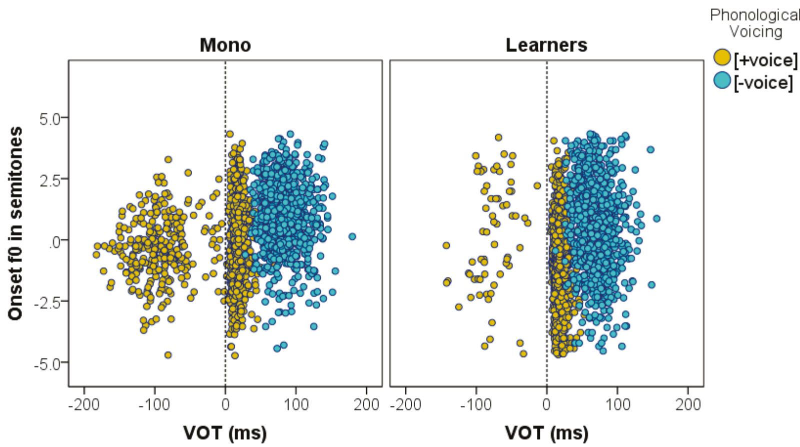


Figure 9. Scatterplots of onset f0 and VOT values in English productions of learners and monolingual speakers.

Specifically, VOT of learners' voiceless stops was shortened, moving towards their own Russian productions. Learners' voiced stops, in contrast, tended towards divergence: not in terms of VOT values (VOT of [+voice] stops, including duration of prevoicing, was not affected) but in terms of prevoicing frequency. Learners produced significantly fewer prevoiced stops in English than monolingual controls. Combined with the fact that learners' Russian prevoiced stops were marked by extra-low onset f0, these findings suggest that learners were targeting prevoiced stops as a distinct non-native category.

There was some evidence for the role of elicitation order: A decrease in the number of prevoiced stops in English was amplified if Russian elicitation occurred immediately prior.

3.1.3. Individual Variability in Drift

The initial stop measures, which showed evidence of L1 drift, included positive VOT of voiceless stops and the frequency of prevoicing in voiced stops. Therefore, we focused on these parameters in evaluating individual drift and its covariation with subjective and objective measures of Russian pronunciation proficiency.

To estimate the magnitude of L1 drift in each participant, we subtracted each learner's mean English voiceless VOT from the grand mean voiceless VOT of all monolingual participants. The resulting value represented how much each learner deviated in their voiceless VOT from the average monolingual norm (greater values represent greater deviation from English monolingual long lag norms and greater approximation of Russian short lag norms).

These values were checked for correlation with each learner's average Russian voiceless VOT. The prediction tested is that learners who were more successful in shortening their Russian voiceless VOT are also expected show a greater amount of drift in their English voiceless VOTs.

The results of the two-tailed Pearson correlation analysis showed a significant negative correlation between the individual drift and individual Russian voiceless VOT ($r = -0.613$, $p = 0.005$). Figure 10 shows that participants with the shortest Russian voiceless VOTs demonstrated the greatest L1 drift in the direction of Russian short lag norms.

We also checked the magnitude of drift parameter for correlations with self-estimated Russian speaking proficiency and self-estimated accentedness (subjective measures), where proficiency and accentedness scores on a seven-point scale were treated as continuous parameters, but neither revealed significant co-dependencies. A similar analysis was conducted for individual frequency of prevoicing across learners' two languages, with no significant results.

Interim summary: Correlation analyses of initial stop data indicated that for VOT of voiceless stops, the magnitude of individual L1 drift was significantly correlated with pronunciation proficiency in Russian, when the latter was measured objectively via acoustic analysis.

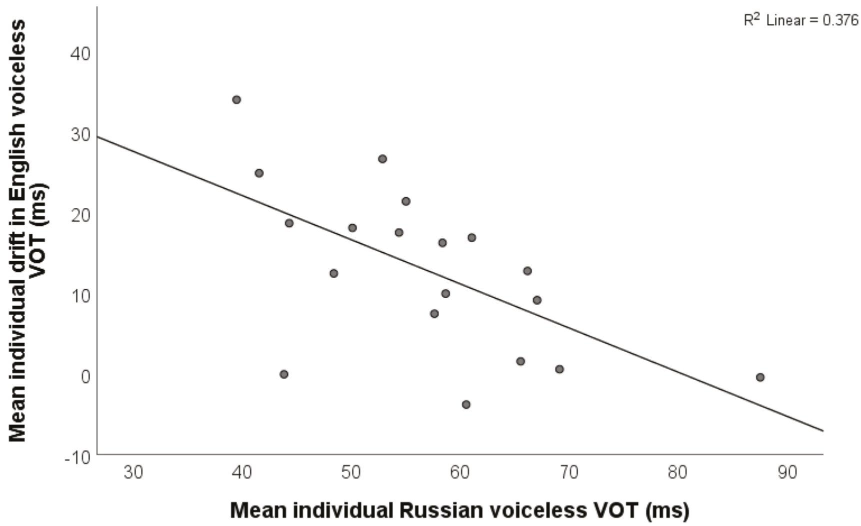


Figure 10. Correlation between individual L1 drift in English voiceless VOT (*y*-axis; larger values indicate greater drift towards Russian-like short lag VOT) and individual Russian voiceless VOT (*x*-axis).

3.2. Final Stops

3.2.1. Learners' Russian vs. English

Similar to the analysis of initial stops, the examination of the acoustic correlates of final voicing assessed whether and to what extent the learners approximated the goal of neutralizing the voicing distinction in word-final position in Russian, while maintaining the contrast in English.

Vowel duration: Duration of the vowel preceding final obstruents was analyzed using an LMM with Language, Voicing, and Language by Voicing as fixed factors. The results demonstrated a significant effect of Voicing, $F(1, 90.19) = 10.01$, $p = 0.002$, and a significant Language by Voicing interaction, $F(1, 90.18) = 13.20$, $p < 0.001$. Vowels were significantly longer before voiced (M = 226 ms, SD = 67 ms) than before voiceless obstruents (M = 159 ms, SD = 52 ms). The difference was greater in learners' English than in their Russian (see Figure 11), explaining the Language by Voicing interaction, and indicating that learners were approximating a reduction in the voicing distinction in Russian, by shortening vowels in the voiced context. While this modification is consistent with partial devoicing of [+voice] final obstruents in Russian, complete neutralization was clearly not achieved.

Constriction duration: Stop closure duration and fricative frication duration were analyzed in two separate LMMs with Language, Voicing, and Language by Voicing as fixed factors. Affricates were absent from this analysis because none were included among the Russian stimuli.

The results for closure duration demonstrated a significant effect of Voicing, $F(1, 69.67) = 411.59$, $p < 0.001$, Language, $F(1, 75.94) = 61.85$, $p < 0.001$, and a significant Language by Voicing interaction, $F(1, 69.67) = 6.05$, $p = 0.016$. Stop closure was significantly longer in voiceless (M = 119 ms, SD = 41 ms) than in voiced stops (M = 76 ms, SD = 28 ms). Closures were also significantly longer in learners' Russian (M = 104 ms, SD = 40 ms) than in their English (M = 94 ms, SD = 41 ms). Finally, as Figure 12 demonstrates, the difference between voiced and voiceless closures was smaller in Russian than in

English, at the expense of longer [+voice] closures in Russian, explaining the significant Language by Voicing interaction. Again, longer [+voice] Russian closures are compatible with partial devoicing of those stops in participants' Russian speech.

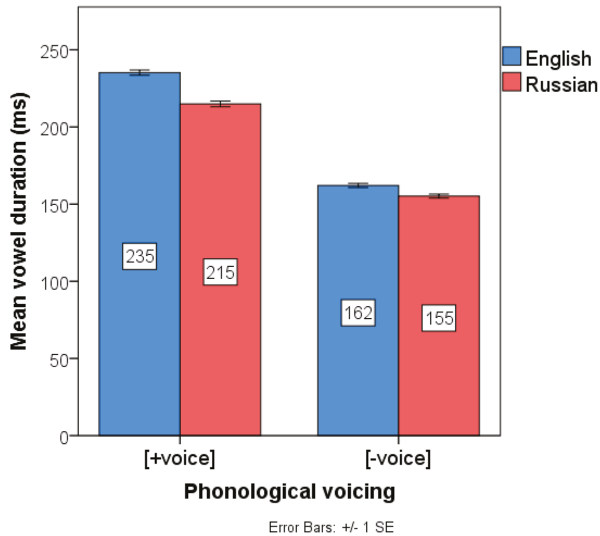


Figure 11. Mean vowel duration before voiced and voiceless final obstruents in learners' English and Russian.

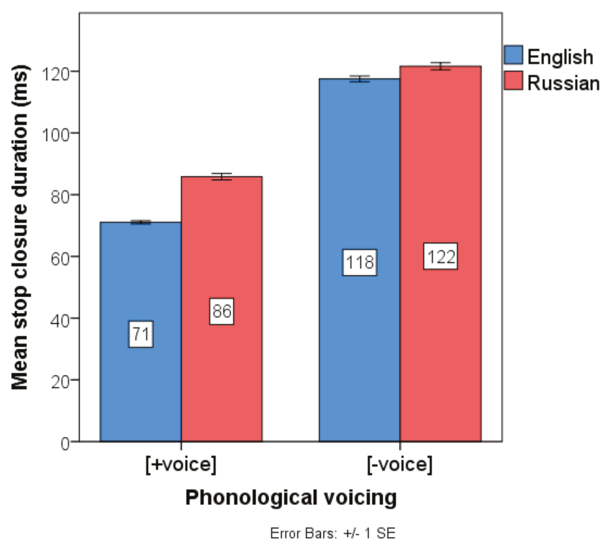


Figure 12. Mean voiced and voiceless final stop closure duration in learners' English and Russian.

Results for frication duration showed no significant effects beyond that of Voicing, $F(1, 18.99) = 35.36$, $p < 0.001$: voiceless fricatives were significantly longer (M = 273 ms, SD = 60 ms) than voiced ones (M = 210 ms, SD = 58 ms).

Voicing duration: The duration of the closure or frication portion of the final obstruent characterized by laryngeal voicing (glottal pulsing), in ms, was submitted to an LMM analysis with Language, Voicing, and Language by Voicing as fixed factors. The results demonstrated a significant effect of Language, F(1, 95.53) = 7.28, p = 0.008, and Voicing, F(1, 99.99) = 678.34, p < 0.001, and a significant Language by Voicing interaction, F(1, 99.99) = 6.47, p = 0.013. English final obstruents contained more voicing (M = 48 ms, SD = 59 ms) than Russian final obstruents (M = 36 ms, SD = 48 ms). Voiced final obstruents had more voicing (M = 83 ms, SD = 55 ms) than voiceless ones (M = 6 ms, SD = 16 ms). The interaction was due to the fact that the difference between voiced and voiceless obstruents was greater in participants' English than in their Russian speech, as shown in Figure 13. The voicing contrast between final obstruents in Russian was reduced via partial devoicing of the voiced obstruents.

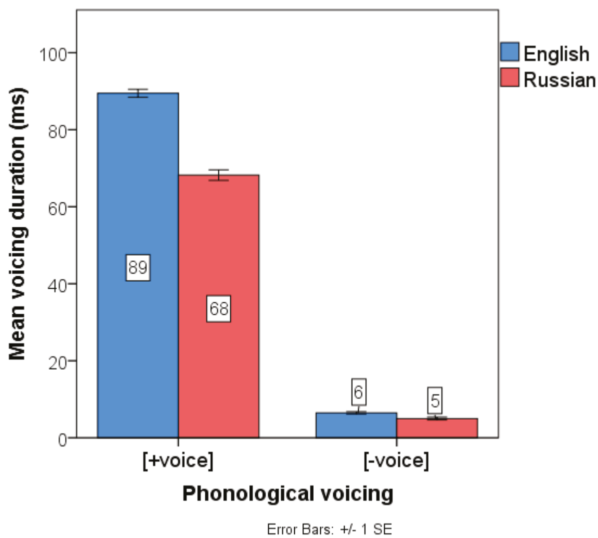


Figure 13. Mean duration of laryngeal voicing during closure or frication portion of the final voiced and voiceless obstruents in learners' English and Russian.

Interim summary: The results for final obstruents overall demonstrated that learners attempted to implement voicing neutralization in their Russian productions. The preceding vowel duration, stop closure duration, and voicing during constriction demonstrated subphonemic but statistically significant tendencies towards partial devoicing of final [+voice] obstruents, although complete neutralization was not achieved. Thus, similarly to initial obstruents, we observed that participants aimed to implement appropriate phonetic differences between English and Russian final obstruents.

3.2.2. Learners' English vs. Monolingual Controls' English

Similar to initial stops, the goal of the analysis was to determine whether learners' English productions were affected by exposure to Russian. If significant differences are detected between experimental and control groups in the acoustic implementation of final obstruents, these could suggest a 'drift' towards Russian norms.

Figure 14 shows the distribution of voiced and voiceless final obstruents, based on the constriction duration (closure duration for stops and frication duration for fricatives) and preceding vowel duration, in the English speech of monolingual controls and learners of Russian. Although this display does not contain all datapoints (affricates were excluded due to extra-long constrictions and a relatively small number of tokens), it provides a representative picture of the differences between monolingual speakers and learners. First, the learners' distribution demonstrates a greater amount of variability in terms of constriction duration. However, most importantly, learners' data demonstrate visibly less separation between the two voicing categories in both dimensions. This suggests that learners' English is drifting towards a reduction in the final voicing contrast, at least in these two acoustic dimensions. The direction of the drift is consistent with an influence of Russian final voicing neutralization. To determine whether these tendencies were statistically significant, each dimension of contrast was subjected to statistical analysis.

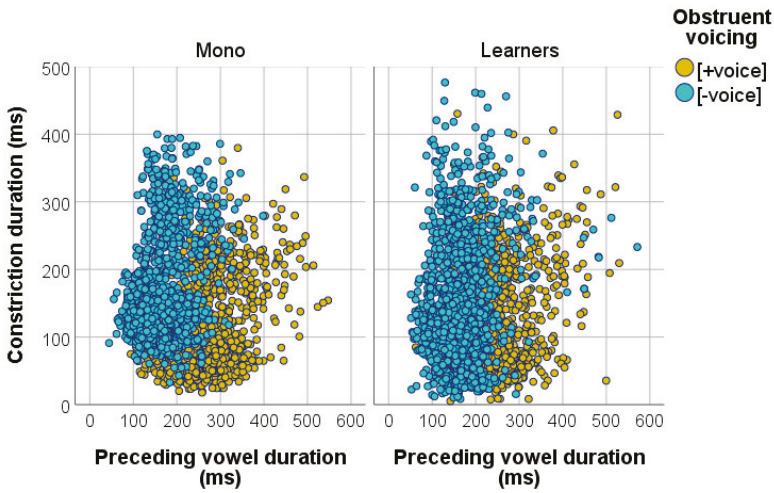


Figure 14. Scatterplot of constriction duration and preceding vowel duration for final stops and fricatives in English speech of monolingual controls and learners.

Vowel duration: Vowel duration was analyzed in an LMM with Group (Learners vs. Controls), Voicing, and Group by Voicing as fixed factors. The results demonstrated a significant effect of Group, $F(1, 36.01) = 4.24$, $p = 0.047$, Voicing, $F(1, 802.35) = 8.51$, $p = 0.004$, and a significant Group by Voicing interaction, $F(1, 5961.38) = 140.31$, $p < 0.001$. Vowels were significantly longer before voiced (M = 250 ms, SD = 68 ms) than before voiceless obstruents (M = 166 ms, SD = 51 ms). Monolingual speakers produced longer vowels (M = 219 ms, SD = 75 ms) than learners (M = 199 ms, SD = 70 ms). The interaction revealed that the vowel duration difference as a function of consonant voicing was greater for monolingual speakers than for learners. This suggests that learners experienced drift towards Russian norms evidenced by partial devoicing of [+voice] obstruents in terms of preceding vowel duration (see Figure 15).

To address the role of elicitation order in the emergence of L1 drift, we conducted an LMM analysis of vowel duration within the group of learners with fixed factors of Voicing, Order (Russian-first or English-first), and Voicing by Order. The results confirmed a significant effect of Voicing, $F(1, 641.98) = 16.09$, $p < 0.001$, but showed no other significant effects.

Constriction duration: Duration of closure (for stop consonants and affricates) and duration of frication (for fricatives and affricates) were analyzed in two separate LMM analyses with Group (Monolinguals vs. Learners), Voicing, and Group by Voicing as fixed factors.

Analysis of closure duration demonstrated a significant effect of Voicing, $F(1, 40.98) = 28.08$, $p < 0.001$, Group, $F(1, 35.82) = 9.26$, $p = 0.004$, and a significant Group by Voicing interaction, $F(1, 4457.49) = 82.83$, $p < 0.001$. Voiceless stops and affricates had significantly longer closures (M = 125 ms, SD = 56 ms) than voiced ones (M = 80 ms, SD = 44 ms), but this difference was considerably more pronounced in monolinguals' than in learners' English, explaining the interaction. As Figure 16 shows, the average difference in closure duration between voiced and voiceless categories was smaller in learners' than in monolinguals' English, and was more comparable to the amount of contrast realized in learners' Russian speech. The reduction in contrast in learners' speech occurred by shortening voiceless closures.

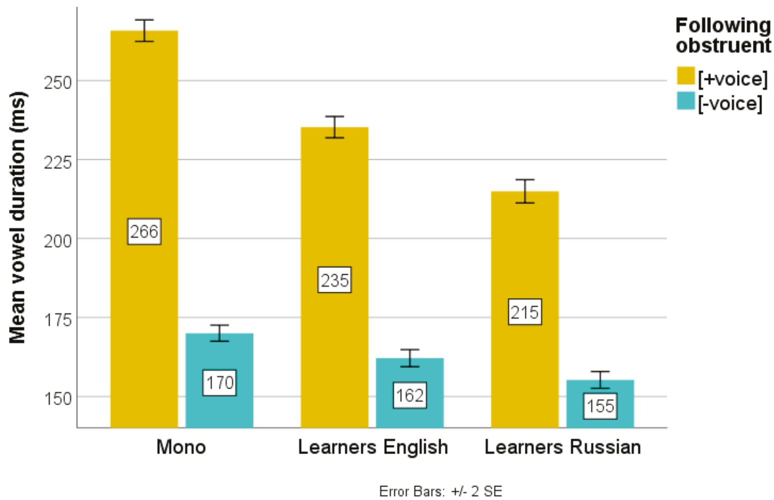


Figure 15. Mean vowel duration before voiced and voiceless obstruents in monolingual controls' and learners' English; learners' Russian productions are provided for comparison.

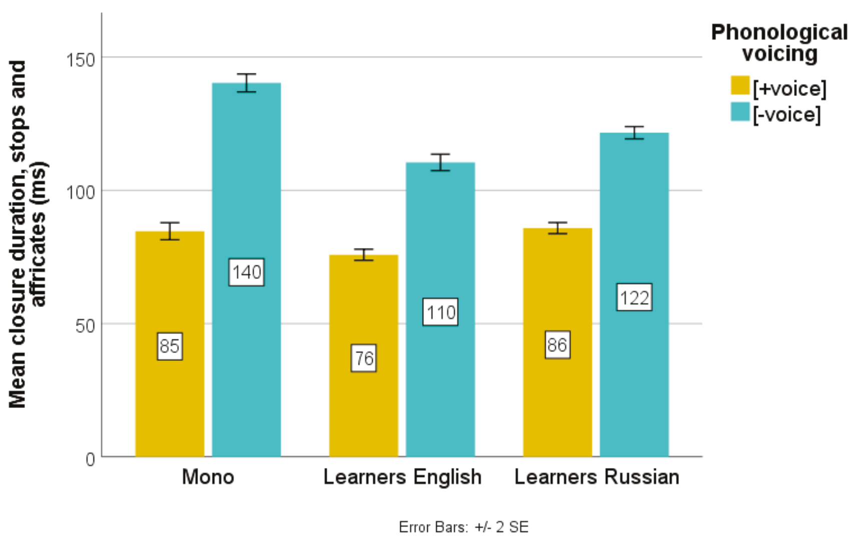


Figure 16. Mean voiced and voiceless closure for final stops and affricates in monolingual controls' and learners' English productions; learners' Russian productions are provided for comparison.

The Order analysis for closure duration revealed no effects beyond the expected effect of Voicing, F(1, 38.01) = 68.86, $p < 0.001$.

Analysis of frication duration demonstrated a significant effect of Voicing, F(1, 65.38) = 9.80, $p = 0.003$, and a significant Group by Voicing interaction, F(1, 1756.68) = 32.18, $p < 0.001$. Voiceless fricatives and affricates presented significantly longer frication duration (M = 257 ms, SD = 62 ms) than voiced ones (M = 198 ms, SD = 54 ms) and the difference was more pronounced in monolinguals' than in learners' English (Figure 17).

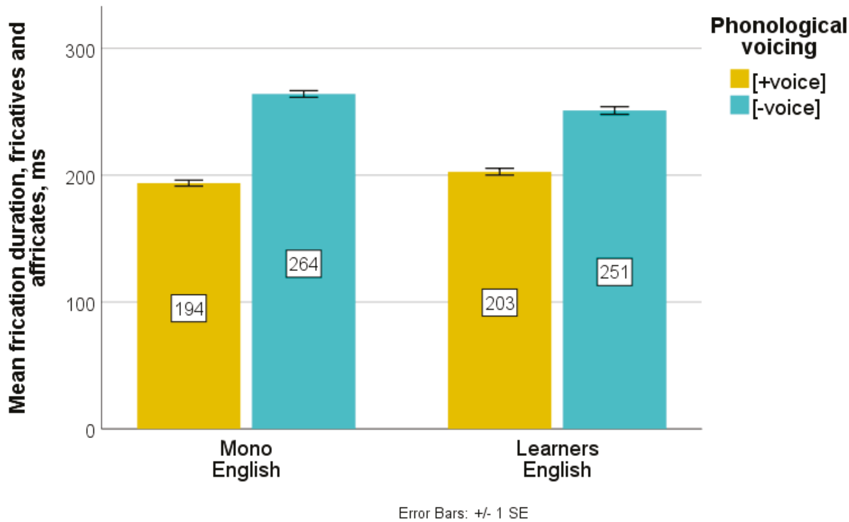


Figure 17. Mean voiced and voiceless frication for final fricatives and affricates in monolingual controls' and learners' English productions.

The Order analysis showed no significant effects of Voicing, Order, or Voicing by Order. This result indicates that the order of language presentation did not affect the magnitude of drift in learners' L1 for fricative duration. The absence of significant Voicing effect also suggests that voicing-dependent differences in final frication duration could be completely neutralized in learners' English speech.

Voicing duration: The duration of glottal pulsing during the frication or closure portion of stops, affricates, and fricatives was analyzed in an LMM with Group, Voicing, and Group by Voicing as fixed factors. The results showed an effect of Voicing, $F(1, 57.38) = 321.56$, $p < 0.001$, and a Group by Voicing interaction, $F(1, 5970.59) = 19.89$, $p < 0.001$. Significantly more laryngeal voicing was detected during the constriction portion of phonologically voiced (M = 90 ms, SD = 57 ms) than voiceless (M = 7, SD = 17) final obstruents. This difference was more pronounced in the speech of monolingual participants than learners, explaining the interaction. As Figure 18 demonstrates, less laryngeal voicing was found in the learners' final [+voice] obstruents than in those of monolinguals, although the reduction was not quite as great as in learners' Russian speech.

The Order analysis for voicing duration confirmed the effect of Voicing, $F(1, 57.38) = 321.56$, $p < 0.001$, and revealed a Voicing by Order interaction, $F(1, 5970.59) = 19.89$, $p < 0.001$, which, unexpectedly, was due to greater neutralization of the final voicing contrast in the English-first condition.

Interim summary: The results demonstrated that learners' English has drifted towards Russian norms of final voicing neutralization in terms of preceding vowel duration, closure duration, frication duration, and voicing during constriction of the final obstruents. In most cases, the magnitude of contrast was reduced compared to monolingual controls but the contrast itself was nevertheless maintained. Only in the case of frication duration did the contrast between voiced and voiceless consonants approach complete neutralization in the English speech of learners.

The effects of order of language elicitation were few and inconsistent. In the case of preceding vowel duration, eliciting Russian first has increased the drift effect in English, while in the case of voicing during constriction, eliciting English first led to a greater drift effect in English.

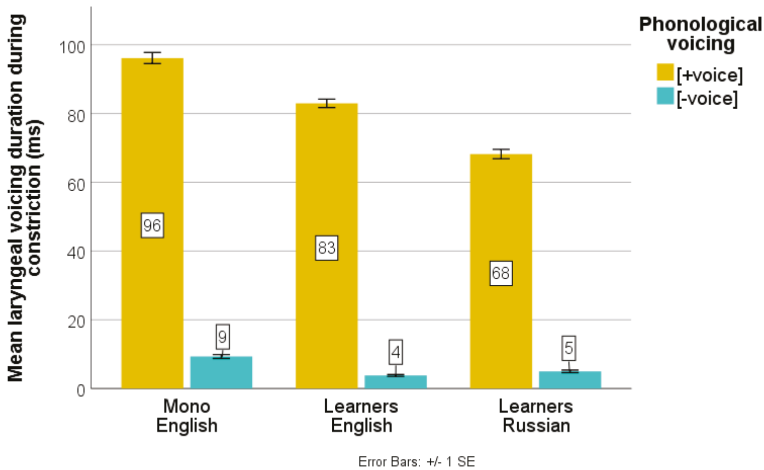


Figure 18. Mean duration of voicing during constriction of final obstruents in monolingual controls' and learners' English; learners' Russian is provided for comparison.

3.2.3. Individual Variability and Drift

The final obstruent measures which showed evidence of drift included preceding vowel duration, closure duration, frication duration, and duration of voicing during closure/frication. These parameters were evaluated for correlations between degree of individual drift and Russian pronunciation proficiency.

Russian pronunciation proficiency was evaluated objectively, as an individual acoustic difference between average voiced and average voiceless productions, and subjectively, as a self-reported speaking proficiency score and accentedness score.

The individual degree of drift was estimated by subtracting, for each participant, the average voiced-voiceless difference from the grand average monolingual voiced-voiceless difference (thus obtaining a 'difference of differences'). Similar calculations were conducted for vowel duration, closure duration, frication duration, and duration of voicing during closure/frication (in all calculations of the voiced-voiceless difference, a smaller value was always subtracted from a larger one, e.g., vowel duration before voiceless obstruents was subtracted from vowel duration before voiced obstruents but voiced closure was subtracted from the voiceless closure).

Calculated in this manner, the individual values for drift were larger if a participant produced a smaller amount of durational difference in a given parameter between the voiced and voiceless obstruent, i.e., if they drifted more in the direction of Russian final voicing neutralization, as compared to an average monolingual difference.

Two-tailed Pearson correlational analyses revealed a significant negative relationship between individual drift in vowel duration ($r = -0.539$, $p = 0.017$), closure duration ($r = -0.666$, $p = 0.002$), and voicing duration ($r = -0.463$, $p = 0.046$) and the voiced-voiceless difference in these parameters in their Russian speech. In other words, those participants who drifted the most towards Russian neutralization norms in their English were also the ones who neutralized the same parameter more successfully in their Russian speech. Figure 19 illustrates the relationship for closure duration. Interestingly, in all three analyses, it was always a subset of the four learners (# 6, 21, 22, and 23) who had lived in Russia who consistently demonstrated the strongest drift, the greatest pronunciation gains, or both (e.g., Figure 19).

The amount of drift in all of these parameters, except voicing duration, also correlated significantly with self-estimated Russian speaking proficiency ($r = 0.526$, $p = 0.017$ for vowel duration and $r = 0.514$, $p = 0.020$ for closure duration), such that participants with high self-estimated Russian proficiency were the ones who drifted towards Russian the most in their English. No measure of drift correlated with the self-estimated degree of accentedness in Russian.

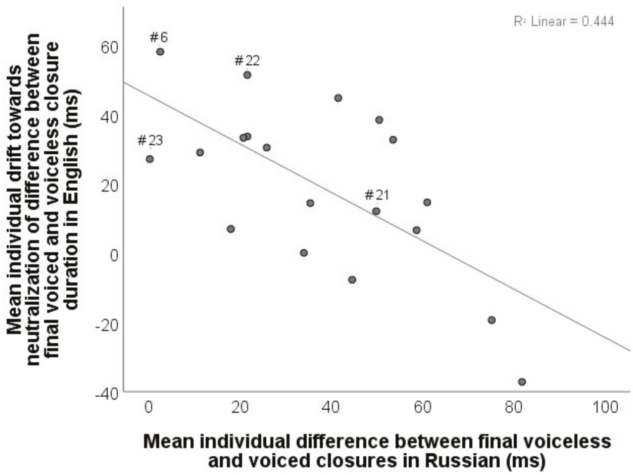


Figure 19. Correlation between average individual L1 drift towards the reduction in voiceless-voiced closure duration difference in final stops/affricates and average individual difference between final voiceless and voiced closures in Russian (y-axis: larger values indicate greater drift towards Russian-like contrast neutralization; x-axis: larger values indicate greater distinction between voiced and voiceless obstruents). Labelled datapoints are participants who had lived in Russia.

Interim summary: For three acoustic correlates of the final voiced-voiceless distinction, correlational analyses revealed a significant dependency between the amount of L1 drift and the degree of pronunciation proficiency in L2, as evaluated objectively via acoustic measures. In two of these cases, a subjective measure of pronunciation proficiency, namely self-evaluated speaking proficiency in Russian along a one- to seven-point Likert scale, also correlated with the magnitude of drift, suggesting that, at least in some cases, such simple self-reported data align with the objective acoustic measures.

3.3. Summary of the Results

The comparisons between Russian and English productions of L2 learners demonstrated that learners, as a group, were attempting to implement the phonetic differences between the two languages (as they pertain to the implementation of voicing in particular) in their Russian productions. At the same time, while producing acoustically distinct targets in their two languages, learners' productions were still quite distinct from the Russian norms.

Despite the relatively modest gains in Russian pronunciation abilities, learners also demonstrated discernable effects of exposure to Russian in their English speech. Comparisons with a monolingual control group revealed significant differences in the majority of examined acoustic parameters. Most differences were compatible with the effect of Russian in the direction of convergence. Additionally, the individual magnitude of L1 drift in the direction of Russian norms was, in many cases, correlated with the degree of pronunciation gains in Russian speech: participants who could be considered more proficient speakers of Russian (as evaluated using acoustic measurements or self-reported speaking proficiency scores) demonstrated the greatest degrees of L1 drift.

4. Discussion

The present study examined native and second language speech of American learners of Russian in order to determine whether classroom exposure to L2 can lead to phonetic changes in learners' L1.

First, to confirm that classroom exposure to L2 resulted in phonetic learning, as evidenced by effective separation of L1 and L2 systems in the speech of the participants, and to evaluate the degree of this learning, we conducted a comparison between the acoustics of learners' Russian and English sounds. The results indicated that, for almost every measure taken, learners produced statistically distinct values in Russian and English. In Russian, their word-initial voiceless stops had shorter VOTs, their word-initial voiced stops had longer prevoicing, the frequency of prevoiced stops was higher, and prevoiced stops were characterized by extra-low onset f0, compared to English. Learners' word-final voiced obstruents were also partially devoiced in Russian, compared to English. All of these modifications were in the direction of approximating native Russian norms: short lag voiceless stops, robustly and near-exclusively prevoiced voiced stops, lower onset f0 after prevoiced stops, and word-final devoicing.

The fact that distinct productions were obtained across learners' L1 and L2 indicates that even at these relatively early stages of learning, taking place while immersed in L1, participants grasped the phonetic differences between similar phones across the languages and were attempting to implement them in their L2 speech. These results fit in with a wide array of similar findings for bilinguals and L2 learners, demonstrating not merged but distinct productions across the two languages (Baker and Trofimovich 2005; Flege and Eefting 1987a, 1987b; Fowler et al. 2008).

At the same time, it is clear that these learners were not near-native like in their L2 pronunciation by any measure. Their initial [−voice] stops were too aspirated and their initial [+voice] category was still dominated by short lag productions, instead of prevoiced ones. Their final [+voice] obstruents were also only slightly devoiced as opposed to fully devoiced, as expected in Russian speech. Therefore, using these acoustic measures, we can conclude that, at least with respect to pronunciation, these learners were not highly proficient/advanced in their L2.

The question remains whether we can expect L2-driven phonetic changes in the learners' native speech for these non-advanced speakers immersed in the L1. The answer given by the present results is 'yes'. A comparison of learners' English to the native English monolingual group revealed acoustically subtle but statistically significant differences, all compatible with the influence of Russian. Learners' initial [−voice] stops were characterized by shortened VOTs, indicating assimilation with Russian. Comparison of learners of Russian with the monolingual group also revealed a tendency to reduce the magnitude of the phonetic contrast between final voiced and voiceless obstruents. This reduction was implemented via partial devoicing of [+voice] final obstruents and is compatible with the effect of the Russian final devoicing process.

This finding suggest that L2 phonological rules can trigger phonetic changes in speakers' L1. The present outcome may also have been helped by the fact that American English is already gravitating towards variable final devoicing, most strongly in fricatives (Davidson 2016), thus facilitating this particular back-transfer from Russian. It is interesting that there was no significant frication duration difference between voiced and voiceless fricatives and affricates in learners' English. Thus, if learners have 'drifted' all the way into Russian-like complete neutralization (with respect to this parameter) then it was only in segments that are especially prone to devoicing in their L1. This finding warrants further research in order to learn more about the conditions under which phonological processes may seep from bilinguals' L2 into their L1.

Interestingly, English [+voice] stops were not affected, neither short lag nor prevoiced ones. Only the frequency of prevoicing showed a change, notably, in the direction of dissimilation from Russian. A similar effect was reported by Huffman and Schuhmann (2016), who indicated that some American learners of Spanish produced fewer or no prevoiced stops in their English after 6 weeks of classroom Spanish instruction. This result merits attention because it demonstrates that L1 phonetic changes in the direction of dissimilation from L2 are possible even at the beginning stages of L2 acquisition—a possibility not provided for in SLM, which predicts that only advanced learners will dissimilate after having created separate categories for L2 sounds. Moreover, this finding indicates that the dissimilatory or assimilatory direction of crosslinguistic interaction can be determined not only by the stage of L2 acquisition but also by the sound category itself. Specifically, the present data suggest that L1 may tend towards dissimilation with L2 when L1 offers a choice of sub-phonemic variants of the same category only one of which is used in L2 to represent the same phonological category.

Overall, phonetic parameters affected by L1 drift were a subset of those used by participants to differentiate their L1 and L2, suggesting that L2 phonetic learning is a natural precursor of L1 drift.

The present evidence of L2-to-L1 phonetic effects in American learners of Russian indicates that even relatively dissimilar language pairings are subject to such phonetic interactions. Assimilatory changes in the acoustics of English obstruents suggest that despite great linguistic differences between the two languages overall and the use of different orthographic symbols for these sounds, Russian and English segments influenced each other. Orthography has been shown to play a powerful role in adult second language learning, which relies greatly on literacy and orthographic input, unlike first language acquisition (Bassetti and Atkinson 2015; Bassetti et al. 2015; Hayes-Harb et al. 2018). Nevertheless, the present study shows, in agreement with previous research on dissimilar language pairs such as English and Korean, that orthographic differences are not insurmountable obstacles for equivalence classification even in highly literate adult language learners. Equivalence classification between English and Russian obstruents, leading to cross-language assimilatory changes, could be facilitated by similarities in the phonological functioning of these segments in the respective languages. The two languages have similar sets of voiced and voiceless obstruents, which contrast initially and intervocalically but assimilate in voicing when in clusters and devoice, to different degrees, when in final position. Additionally, equivalent phonological categories across these two languages can be realized in phonetically identical ways, albeit in different contextual environments, e.g., English [−voice] stops can be implemented with short lag VOT word-medially before unstressed vowels, similarly to Russian [−voice] stops.

The fact that L2-to-L1 phonetic effects were detected for traditional classroom learners indicates that L2 immersion is not a necessary condition and that the amount of L2 experience and exposure received via classroom instruction can be sufficient to trigger such changes. Our participants did produce an acoustic distinction between L1 and L2 obstruents as a group, which suggest that this degree of phonetic learning may be required for L1 drift to initiate. Conversely, advanced pronunciation proficiency is most likely not a pre-requisite for L1 drift. Nevertheless, prior L2-immersion, which was not ongoing at the time of participation and was limited to 2-4 months, appeared to intensify the degree of L1 drift for these participants in some measures, in parallel with improving the authenticity of their L2 pronunciation.

Moreover, these findings strongly suggest that a significant reduction in the use of L1 is also not a necessary condition for L1 phonetic drift. It is very unlikely that participants in the present study experienced a substantial reduction in the amount or quality of L1 use as a result of studying Russian at the university. It is also unlikely that they were exposed to much spoken Russian through overhearing (concomitantly, also reducing the 'overhearing' of English), the way immersed learners are. Thus, even in the absence of considerable reduction in L1 use, first language can and does drift towards the phonetics of L2 in comparable sound categories. To complement this finding, other work demonstrated that even in bilingual or immigrant settings, where L1 use reduction is likely, L1 use does not always correlate with the quality of L1 pronunciation (Guion et al. 2000; Hopp and Schmid 2013). The diversity of results with regard to the effect of L1 use on L1 speech indicates that its role is not fully understood and merits further attention.

Additionally, L1 drift in the present work was not the result of an immediate 'spill-over' effect from one language to another. The order of Russian and English elicitation was counterbalanced, and the analysis of elicitation order effects showed that order did not condition the presence of the L1 drift, although it could sometimes increase its magnitude.

Although it is unlikely that learners' L1 use was substantially decreased by their enrollment in Russian courses, another possibility is that L1 inhibition, not L1 use reduction, is what paved the road for L2-to-L1 phonetic effects. Some authors have argued that successful L1 inhibition is important for effective L2 learning. For example, Linck et al. (2009) showed that immersed L2 learners fared better in acquired L2 proficiency but worse in L1 lexical retrieval than classroom learners, and argued that greater L1 inhibition in immersion settings was responsible. Moreover, Levy et al. (2007) demonstrated that even a short laboratory training session was enough to trigger L1 inhibition effects in lexical retrieval. This suggests that a relatively limited time of classroom L2 learning may also be sufficient to trigger L1 inhibition and, therefore, L1 drift. Furthermore, if laboratory training can induce L1 inhibition and L1 inhibition can trigger L1 drift, we could expect L2-to-L1 phonetic effects under laboratory conditions. This is precisely what was demonstrated by Kartushina et al. (2016a) who showed drift in L1 vowels towards similar non-native ones after short-term visual articulatory feedback training. Nevertheless, further research is necessary to fully understand the role of L1 inhibition in the susceptibility of the L1 sound system to the influence of the L2.

A related issue is the longevity of L1 drift observed in laboratory conditions, after a short L2 immersion, or in the course of classroom L2 acquisition. It is rather plausible that such effects may be short-lived. In fact, Kartushina and Martin (2019) showed that, four months after studying abroad, participants experienced a 'return drift' towards native-like phonetics in their L1 (see also Chang (2019b) for similar findings). It is possible that our learners would lose the effects of Russian on their L1 if or when they discontinue their Russian studies. Such short-term phonetic changes in L1 may be qualitatively different from language attrition, which is believed to develop over longer periods of time and have greater strength of 'inertia' in resisting the 'rebound' back towards native-like values when speakers return to L1 immersion and no longer actively use their L2 (Chang 2019a). Additionally, the 'novelty' effect may boost the cross-language drift at the early stages of language acquisition (Chang 2013). Ultimately, the observation that the L1 can respond flexibly and intricately to the changing circumstances of learners' language use and environment demonstrates its great plasticity and adaptability and argues against maturation-related limits on phonetic learning.

Finally, there are a number of factors we could not address in the present study, but which merit serious attention in future research. Among those is the role of exposure to a non-native like L1 in triggering L1 phonetic drift as well as factors such as motivation for learning and attitudes towards the L2 and its associated culture.

An implicit assumption in previous research examining L2-to-L1 phonetic effects has been that it is the exposure to and use of L2, per se, that triggers L1 phonetic drift. This assumption is supported in the current study by the observation that L1 drift co-occurred with a degree of phonetic learning in L2 sufficient to produce two distinct phonetic systems. However, under many scenarios of L2 learning and use, learners are also exposed, to varying degrees, to a non-native-like and accented L1. In the present case, all instructors in the Russian courses attended by our learners were native speakers of Russian. An informal survey of Russian instructors at Purdue University indicated that during class students may be addressed in English anywhere between 15% and 70% of the time, depending on the course level and individual proficiencies of class participants. This suggests that, especially at the beginner level of Russian instruction, learners are exposed to a considerable amount of Russian-accented English speech. Some acoustic characteristics of Russian-accented English would be very similar to the ones observed in native English that has drifted towards Russian (e.g., shorter voiceless VOTs of initial stops, partial or complete devoicing of final obstruents).

At present, we know very little about the role that such exposure may play in the development of apparent L1 drift. Nevertheless, some research suggests that L2-accented L1 input may contribute to non-native-like L1 productions in bilinguals. For example, Mora and Nadeu (2012) and Mora et al. (2015) suggest that the partial merging of Catalan /e/ and /ɛ/ vowels in the speech of their participants could be due, in part, to their exposure to Spanish-accented Catalan (see also Sebastián-Gallés et al. 2005). If L1 drift in classroom language learners is guided, in part or primarily by the Russian-accented English input provided by their instructors, drift may develop as L1 phonetic accommodation to the instructor. In this case, it could be impacted by factors shown to mediate accommodation, such as considerations of dominance, prestige, and speakers' disposition and attitudes towards each other, and language distance (Babel 2012; Babel et al. 2014; Kim et al. 2011; Pardo 2006; Pardo et al. 2012; Pardo et al. 2013; Yu et al. 2013).

Additionally, learners' motivations for studying the chosen language and their global attitudes towards the associated culture and native speakers of their L2 could further mediate the propensity to drift. Previous research has shown that language attitudes and considerations of prestige can influence cross-language interaction in bilinguals (Gatbonton et al. 2005; Gatbonton et al. 2011; Giles et al. 1977; Law et al. 2019), but considerably more work is needed to determine the role of such factors in L2-to-L1 phonetic effects.

5. Conclusions

The study presented here examined the acoustic characteristics of word-initial and word-final obstruents in English and Russian speech of Americans studying Russian as a second language in the traditional college classroom. The results demonstrated that learners' native English productions were acoustically distinct from those of monolingual speakers of American English, primarily in the direction of assimilation to Russian acoustics. We interpret these results to indicate that advanced mastery of another language or extensive long-term exposure to L2 are not necessary for the phonetics of a second language to affect the first language of learners. The current results also suggest that a reduction in the use of L1, which typically accompanies L2 immersion, is also not a necessary condition of L1 drift. Overall, the results demonstrate that the acoustics of L1 production can undergo subtle but systematic adjustments as a result of new linguistic experiences such as second language learning in classroom settings.

Author Contributions: Conceptualization, O.D., A.J. and J.A.S.; methodology, O.D.; formal analysis, O.D.; investigation, O.D., A.J. and J.A.S.; data curation, O.D.; writing—original draft preparation, O.D.; writing—review and editing, A.J. and J.A.S.; visualization, O.D.; supervision, O.D.; project administration, O.D., A.J. and J.A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors would like to acknowledge the contributions of undergraduate research assistants at Purdue University: Alyssa Nymeyer, Audrey Bengert, Bethany Sexton, Emilie Watson, and Alexis Tews. We would also like to thank Russian language instructors at Purdue University, Olga Lyanda-Geller and Amina Gabrielova, for assistance in recruiting participants and providing information about teaching practices.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. English minimal pairs with voiced and voiceless plosives in word-initial position.

English Initial Stops					
bad	pad	dank	tank	gap	cap
bat	pat	dab	tab	gab	cab
bond	pond	dot	tot	got	cot
bob	pop	dog	tog	God	cod
bun	pun	duck	tuck	gut	cut
buck	puck	dug	tug	gum	come

Table A2. English minimal pairs with voiced and voiceless obstruents in word-final position.

English Final Obstruents									
Stops						Fricatives		Affricates	
cab	cap	mad	mat	tag	tack	seize	cease	badge	batch
tab	tap	pad	pat	hag	hack	buzz	bus	ridge	rich
mob	mop	hod	hot	hog	hock	fuzz	fuss		
hob	hop	cod	cot	dog	dock	raise	race		
cub	cup	cud	cut	dug	duck	leave	leaf		
pub	pup	mud	mutt	tug	tuck	save	safe		

Table A3. Russian minimal pairs with voiced and voiceless plosives in word-initial position.

Russian Initial Stops					
бар	пар	дом	том	гость	кость
bar	*steam*	*house*	*volume*	*guest*	*bone*
[bar]	[par]	[dom]	[tom]	[gostʲ]	[kostʲ]
борт	порт	док	ток	гол	кол
board	*port*	*dock*	*current*	*goal*	*stake*
[bort]	[port]	[dok]	[tok]	[gol]	[kol]
барс	парк	день	тень	гипс	кисть
leopard	*park*	*day*	*shadow*	*cast*	*brush*
[bars]	[park]	[denʲ]	[tenʲ]	[gips]	[kistʲ]
быль	пыль	душ	тушь	гид	кит
true story	*dust*	*shower*	*ink*	*guide*	*whale*
[bɨlʲ]	[pɨlʲ]	[dush]	[tush]	[git]	[kit]
бас	паз	даль	тальк	год	кот
bass	*groove*	*distance*	*talc*	*year*	*cat*
[bas]	[pas]	[dalʲ]	[talʲk]	[got]	[kot]
боль	пол	диск	тис	гусь	куст
pain	*floor*	*disc*	*yew*	*goose*	*shrub*
[bolʲ]	[pol]	[disk]	[tis]	[gusʲ]	[kust]

Table A4. Russian minimal pairs with voiced and voiceless obstruents in word-final position.

Russian Final Obstruents							
гриб	хрип	пруд	прут	рог	рок	паз	бас
mushroom	*wheeze*	*pond*	*twig*	*horn*	*doom*	*groove*	*bass*
[grip]	[xrip]	[prut]	[prut]	[rok]	[rok]	[pas]	[bas]
зуб	суп	код	кот	маг	мак	воз	нос
tooth	*soup*	*code*	*cat*	*magician*	*poppy*	*cartload*	*nose*
[zup]	[sup]	[kot]	[kot]	[mak]	[mak]	[vos]	[nos]
сноб	сноп	род	рот	луг	лук	приз	рис
snob	*sheaf*	*genus*	*mouth*	*meadow*	*onion*	*prize*	*rice*
[snop]	[snop]	[rot]	[rot]	[luk]	[luk]	[pris]	[ris]
краб	храп	лад	брат	блог	блок	груз	трус
crab	*snort*	*harmony*	*brother*	*blog*	*block*	*load*	*coward*
[krap]	[xrap]	[lat]	[brat]	[blok]	[blok]	[grus]	[trus]
лоб	клоп	след	цвет	бриг	крик	хлев	блеф
forehead	*bedbug*	*footprint*	*color*	*brig*	*scream*	*pigpen*	*bluff*
[lop]	[klop]	[slet]	[t͡svet]	[brik]	[krik]	[xlef]	[blef]
хлеб	креп	гид	кит	бег	век	слив	риф
bread	*crepe*	*guide*	*whale*	*run*	*century*	*sink*	*reef*
[xlep]	[krep]	[git]	[kit]	[bek]	[vek]	[slif]	[rif]

References

Au, Terry Kit-fong, Leah M. Knightly, Sun-Ah Jun, and Janet S. Oh. 2002. Overhearing a Language During Childhood. *Psychological Science* 13: 238–43. [CrossRef] [PubMed]

Babel, Molly. 2012. Evidence for Phonetic and Social Selectivity in Spontaneous Phonetic Imitation. *Journal of Phonetics* 40: 177–89. [CrossRef]

Babel, Molly, Grant McGuire, Sophia Walters, and Alice Nicholls. 2014. Novelty and Social Preference in Phonetic Accommodation. *Laboratory Phonology* 5: 123–50. [CrossRef]

Baker, Wendy, and Pavel Trofimovich. 2005. Interaction of Native- and Second-Language Vowel System(s) in Early and Late Bilinguals. *Language and Speech* 48: 1–27. [CrossRef]

Barlow, Jessica A., Paige E. Branson, and Ignatius S. B. Nip. 2013. Phonetic Equivalence in the Acquisition of /l/ by Spanish–English Bilingual Children. *Bilingualism: Language and Cognition* 16: 68–85. [CrossRef]

Bassetti, Bene, and Nathan Atkinson. 2015. Effects of Orthographic Forms on Pronunciation in Experienced Instructed Second Language Learners. *Applied Psycholinguistics* 36: 67–91. [CrossRef]

Bassetti, Bene, Paola Escudero, and Rachel Hayes-Harb. 2015. Second Language Phonology at the Interface between Acoustic and Orthographic Input. *Applied Psycholinguistics* 36: 1–6. [CrossRef]

Bergmann, Christopher, Amber Nota, Simone A. Sprenger, and Monika S. Schmid. 2016. L2 Immersion Causes Non-Native-like L1 Pronunciation in German Attriters. *Journal of Phonetics* 58: 71–86. [CrossRef]

Boersma, Paul, and David Weenink. 2018. Praat: Doing Phonetics by Computer [Computer Program]. Available online: http://www.fon.hum.uva.nl/praat/ (accessed on 6 April 2016).

Bullock, Barbara E., and Chip Gerfen. 2004. Phonological Convergence in a Contracting Language Variety. *Bilingualism: Language and Cognition* 7: 95–104. [CrossRef]

Caramazza, Alfonso, Grace H. Yeni-Komshian, Edgar B. Zurif, and Ettore Carbone. 1973. The Acquisition of a New Phonological Contrast: The Case of Stop Consonants in French-English Bilinguals. *The Journal of the Acoustical Society of America* 54: 421–28. [CrossRef]

Chang, Charles. B. 2012. Rapid and Multifaceted Effects of Second-Language Learning on First-Language Speech Production. *Journal of Phonetics* 40: 249–268. [CrossRef]

Chang, Charles. B. 2013. A Novelty Effect in Phonetic Drift of the Native Language. *Journal of Phonetics* 41: 520–33. [CrossRef]

Chang, Charles. B. 2019a. Phonetic Drift. In *The Oxford Handbook of Language Attrition*. Edited by Monika S. Schmid and Barbara Köpke. Oxford: Oxford University Press, pp. 191–203.

Chang, Charles. B. 2019b. Language Change and Linguistic Inquiry in a World of Multicompetence: Sustained Phonetic Drift and its Implications for Behavioral Linguistic Research. *Journal of Phonetics* 73: 96–113. [CrossRef]

Davidson, Lisa. 2016. Variability in the Implementation of Voicing in American English Obstruents. *Journal of Phonetics* 54: 35–50. [CrossRef]

Davis, Mark. 2008. The Corpus of Contemporary American English (COCA): One Billion Words, 1990–2019. Available online: https://www.english-corpora.org/coca/ (accessed on 5 May 2015).

De Leeuw, Esther. 2019. Native Speech Plasticity in the German-English Late Bilingual Stefanie Graf: A Longitudinal Study over Four Decades. *Journal of Phonetics* 73: 24–39. [CrossRef]

De Leeuw, Esther, Monika S. Schmid, and Ineke Mennen. 2010. The Effects of Contact on Native Language Pronunciation in an L2 Migrant Setting. *Bilingualism: Language and Cognition* 13: 33–40. [CrossRef]

De Leeuw, Esther, Aurela Tusha, and Monika S. Schmid. 2018. Individual Phonological Attrition in Albanian–English Late Bilinguals. *Bilingualism: Language and Cognition* 21: 278–95. [CrossRef]

Dmitrieva, Olga, Allard Jongman, and Joan Sereno. 2010. Phonological Neutralization by Native and Non-Native Speakers: The Case of Russian Final Devoicing. *Journal of Phonetics* 38: 483–92. [CrossRef]

Dmitrieva, Olga, Fernando Llanos, Amanda A. Shultz, and Alexander L. Francis. 2015. Phonological Status, Not Voice Onset Time, Determines the Acoustic Realization of Onset F0 as a Secondary Voicing Cue in Spanish and English. *Journal of Phonetics* 49: 77–95. [CrossRef]

Flege, James E. 1987. The Production of 'New' and 'Similar' Phones in a Foreign Language: Evidence for the Effect of Equivalence Classification. *Journal of Phonetics* 15: 47–65. [CrossRef]

Flege, James E. 1995. Second Language Speech Learning Theory, Findings, and Problems. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Edited by Winifred Strange. Timonium: York Press, pp. 233–77.

Flege, James E. 2003. Assessing Constraints on Second-Language Segmental Production and Perception James Emil Flege. In *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*. Edited by Antje S. Meyer and Niels O. Schiller. Berlin: Mouton de Gruyter, pp. 319–55.

Flege, James, and Ocke-Schwen Bohn. 2020. The Revised Speech Learning Model. Available online: https://www.researchgate.net/publication/342923241 (accessed on 18 October 2020).

Flege, James E., and Wieke Eefting. 1987a. Production and Perception of English Stops by Native Spanish Speakers. *Journal of Phonetics* 15: 67–83. [CrossRef]

Flege, James E., and Wieke Eefting. 1987b. Cross-Language Switching in Stop Consonant Perception and Production by Dutch Speakers of English. *Speech Communication* 6: 185–202. [CrossRef]

Fowler, Carol A., Valery Sramko, David J. Ostry, Sarah A. Rowland, and Pierre Hallé. 2008. Cross Language Phonetic Influences on the Speech of French-English Bilinguals. *Journal of Phonetics* 36: 649–63. [CrossRef] [PubMed]

Gatbonton, Elizabeth, Pavel Trofimovich, and Michael Magid. 2005. Learners' Ethnic Group Affiliation and L2 Pronunciation Accuracy: A Sociolinguistic Investigation. *TESOL Quarterly* 39: 489–511. [CrossRef]

Gatbonton, Elizabeth, Pavel Trofimovich, and Norman Segalowitz. 2011. Ethnic Group Affiliation and Patterns of Development of a Phonological Variable. *The Modern Language Journal* 95: 188–204. [CrossRef]

Giles, Howard, Donald M. Taylor, and Richard Y. Bourhis. 1977. Dimensions of Welsh Identity. *European Journal of Social Psychology* 7: 165–74. [CrossRef]

Guion, Susan G. 2003. The Vowel Systems of Quichua-Spanish Bilinguals. *Phonetica* 60: 98–128. [CrossRef]

Guion, Susan G., James E. Flege, and Jonathan D. Loftin. 2000. The Effect of L1 Use on Pronunciation in Quichua–Spanish Bilinguals. *Journal of Phonetics* 28: 27–42. [CrossRef]

Harada, Tetsuo. 2003. L2 Influence on L1 Speech in the Production of VOT. Paper presented at the 15th International Congress of Phonetic Sciences, Barcelona, Spain, August 3–9; Edited by Maria-Josep Solé, Daniel Recasens and Joaquín Romero. Barcelona: Causal Productions, pp. 1085–88.

Hayes-Harb, Rachel, Kelsey Brown, and Bruce L. Smith. 2018. Orthographic Input and the Acquisition of German Final Devoicing by Native Speakers of English. *Language and Speech* 61: 547–64. [CrossRef]

Herd, Wendy J., Robin L. Walden, Whitney L. Knight, and Savana N. Alexander. 2015. Phonetic Drift in a First Language Dominant Environment. Paper presented at the Meetings on Acoustics, Jacksonville, FL, USA, November 2–6; Volume 23, pp. 1–9. [CrossRef]

Hopp, Holger, and Monika S. Schmid. 2013. Perceived Foreign Accent in First Language Attrition and Second Language Acquisition: The Impact of Age of Acquisition and Bilingualism. *Applied Psycholinguistics* 34: 361–94. [CrossRef]

Huffman, Marie K., and Katharina S. Schuhmann. 2016. Effect of Early L2 Learning on L1 Stop Voicing. Paper presented at the Meetings on Acoustics, Pittsburgh, PA, USA, September 9; Pittsburgh: Acoustical Society of America, Volume 23, pp. 1–10. [CrossRef]

Kartushina, Natalia, and Clara D. Martin. 2019. Third-Language Learning Affects Bilinguals' Production in Both Their Native Languages: A Longitudinal Study of Dynamic Changes in L1, L2 and L3 Vowel Production. *Journal of Phonetics* 77: 1–21. [CrossRef]

Kartushina, Natalia, Alexis Hervais-Adelman, Ulrich Hans Frauenfelder, and Narly Golestani. 2016a. Mutual Influences between Native and Non-Native Vowels in Production: Evidence from Short-Term Visual Articulatory Feedback Training. *Journal of Phonetics* 57: 21–39. [CrossRef]

Kartushina, Natalia, Ulrich H. Frauenfelder, and Narly Golestani. 2016b. How and When Does the Second Language Influence the Production of Native Speech Sounds: A Literature Review. *Language Learning* 66: 155–86. [CrossRef]

Kim, Midam, William S. Horton, and Ann R. Bradlow. 2011. Phonetic Convergence in Spontaneous Conversations as a Function of Interlocutor Language Distance. *Laboratory Phonology* 2: 125–56. [CrossRef]

Knightly, Leah M., Sun-Ah Jun, Janet S. Oh, and Terry Kit-fong Au. 2003. Production Benefits of Childhood Overhearing. *The Journal of the Acoustical Society of America* 114: 465–74. [CrossRef] [PubMed]

Kulikov, Vladimir. 2012. Voicing and Voice Assimilation in Russian Stops. Ph.D. dissertation, University of Iowa, Iowa City, IA, USA, July. [CrossRef]

Lang, Benjamin, and Lisa Davidson. 2019. Effects of Exposure and Vowel Space Distribution on Phonetic Drift: Evidence from American English Learners of French. *Language and Speech* 62: 30–60. [CrossRef] [PubMed]

Law, Wai Ling, Olga Dmitrieva, and Alexander L. Francis. 2019. Language Attitudes Modulate Phonetic Interactions between Languages in Bilingual Speakers in Diglossic Settings. *Linguistic Approaches to Bilingualism*. [CrossRef]

Lev-Ari, Shiri, and Sharon Peperkamp. 2013. Low Inhibitory Skill Leads to Non-Native Perception and Production in Bilinguals' Native Language. *Journal of Phonetics* 41: 320–31. [CrossRef]

Levy, Benjamin J., Nathan D. McVeigh, Alejandra Marful, and Michael C. Anderson. 2007. Inhibiting Your Native Language: The Role of Retrieval-Induced Forgetting During Second-Language Acquisition. *Psychological Science* 18: 29–34. [CrossRef]

Linck, Jared A., Judith F. Kroll, and Gretchen Sunderman. 2009. Losing Access to the Native Language While Immersed in a Second Language Evidence for the Role of Inhibition in Second-Language Learning. *Psychological Science* 20: 1507–15. [CrossRef]

Lisker, Leigh, and Arthur S. Abramson. 1964. A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *WORD* 20: 384–422. [CrossRef]

Mack, Molly. 1989. Consonant and Vowel Perception and Production: Early English-French Bilinguals and English Monolinguals. *Perception & Psychophysics* 46: 187–200. [CrossRef]

MacLeod, Andrea A. N., and Carol Stoel-Gammon. 2005. Are Bilinguals Different? What VOT Tells Us about Simultaneous Bilinguals. *Journal of Multilingual Communication Disorders* 3: 118–27. [CrossRef]

Major, Roy C. 1992. Losing English as a First Language. *The Modern Language Journal* 76: 190–208. [CrossRef]

Mayr, Robert, Sacha Price, and Ineke Mennen. 2012. First Language Attrition in the Speech of Dutch–English Bilinguals: The Case of Monozygotic Twin Sisters. *Bilingualism: Language and Cognition* 15: 687–700. [CrossRef]

Mora, Joan C., and Marianna Nadeu. 2012. L2 Effects on the Perception and Production of a Native Vowel Contrast in Early Bilinguals. *International Journal of Bilingualism* 16: 484–500. [CrossRef]

Mora, Joan C., James L. Keidel, and James E. Flege. 2015. Effects of Spanish Use on the Production of Catalan Vowels by Early Spanish-Catalan Bilinguals. In *Current Issues in Linguistic Theory*. Edited by Joaquín Romero and María Riera. Amsterdam: John Benjamins, Volume 335, pp. 33–54. [CrossRef]

Pardo, Jennifer S. 2006. On Phonetic Convergence during Conversational Interaction. *The Journal of the Acoustical Society of America* 119: 2382–93. [CrossRef]

Pardo, Jennifer S., Rachel Gibbons, Alexandra Suppes, and Robert M. Krauss. 2012. Phonetic Convergence in College Roommates. *Journal of Phonetics* 40: 190–97. [CrossRef]

Pardo, Jennifer S., Isabel Cajori Jay, Risa Hoshino, Sara Maria Hasbun, Chantal Sowemimo-Coker, and Robert M. Krauss. 2013. Influence of Role-Switching on Phonetic Convergence in Conversation. *Discourse Processes* 50: 276–300. [CrossRef]

Peng, Shu-hui. 1993. Cross-Language Influence on the Production of Mandarin /f/ and /x/ and Taiwanese /h/ by Native Speakers of Taiwanese Amoy. *Phonetica* 50: 245–60. [CrossRef]

Ringen, Catherine, and Vladimir Kulikov. 2012. Voicing in Russian Stops: Cross-Linguistic Implications. *Journal of Slavic Linguistics* 20: 269–86. [CrossRef]

Russian National Corpus. 2003. Available online: https://ruscorpora.ru/old/en/index.html (accessed on 5 May 2015).

Sancier, Michele L., and Carol A. Fowler. 1997. Gestural Drift in a Bilingual Speaker of Brazilian Portuguese and English. *Journal of Phonetics* 25: 421–36. [CrossRef]

Schmid, Monika S. 2013. First Language Attrition. *Linguistic Approaches to Bilingualism* 3: 94–115. [CrossRef]

Schuhmann, Katharina S, and Marie K Huffman. 2015. L1 drift and L2 category formation in second language learning. Paper presented at the 18th International Congress of Phonetic Sciences, Glasgow, UK, August 10–14.

Sebastián-Gallés, Núria, Sagrario Echeverría, and Laura Bosch. 2005. The Influence of Initial Exposure on Lexical Representation: Comparing Early and Simultaneous Bilinguals. *Journal of Memory and Language* 52: 240–55. [CrossRef]

Simonet, Miquel. 2010. Dark and Clear Laterals in Catalan and Spanish: Interaction of Phonetic Categories in Early Bilinguals. *Journal of Phonetics* 38: 663–78. [CrossRef]

Simonet, Miquel, and Mark Amengual. 2020. Increased Language Co-Activation Leads to Enhanced Cross-Linguistic Phonetic Convergence. *International Journal of Bilingualism* 24: 208–21. [CrossRef]

Sundara, Megha, and Shari Baum. 2006. Production of Coronal Stops by Simultaneous Bilingual Adults. *Bilingualism: Language and Cognition* 9: 97–114. [CrossRef]

Tobin, Stephen J., Hosung Nam, and Carol A. Fowler. 2017. Phonetic Drift in Spanish-English Bilinguals: Experiment and a Self-Organizing Model. *Journal of Phonetics* 65: 45–59. [CrossRef]

Ulbrich, Christiane, and Mikhail Ordin. 2014. Can L2-English Influence L1-German? The Case of Post-Vocalic /r/. *Journal of Phonetics* 45: 26–42. [CrossRef]

Yu, Alan C. L., Carissa Abrego-Collier, and Morgan Sonderegger. 2013. Phonetic Imitation from an Individual-Difference Perspective: Subjective Attitude, Personality and 'Autistic' Traits. *PLoS ONE* 8: e74746. [CrossRef]

languages

MDPI

Article

Information-Seeking Question Intonation in Basque Spanish and Its Correlation with Degree of Contact and Language Attitudes

Magdalena Romera [1] and Gorka Elordieta [2,*]

1 Department of Humanities and Education, Public University of Navarre (UPNA), 31006 Pamplona, Spain; magdalena.romera@unavarra.es

2 Department of Linguistics and Basque Studies, University of the Basque Country (UPV/EHU), 01006 Vitoria-Gasteiz, Spain

* Correspondence: gorka.elordieta@ehu.eus

Received: 18 November 2020; Accepted: 9 December 2020; Published: 14 December 2020

Abstract: The present study analyzes the prosodic characteristics of the variety of Spanish in contact with Basque (in the Basque Country, Spain). We focus on information-seeking yes/no questions, which present different intonation contours in Spanish and Basque. In Castilian Spanish, these sentences end in a rising contour, whereas in Basque, they end in a falling or rising–falling circumflex contour. In our previous work, this topic was investigated among the urban populations of Bilbao and San Sebastian. The results were that 79% of information-seeking yes/no questions had final falling intonational configurations. All the speakers presented a substantial presence of final falls regardless of their linguistic profile, but there were differences among speakers in the degree of presence of such features. A correlation was observed between the dependent variable of 'frequency of occurrence of final falls in absolute interrogatives' and social factors, such as 'degree of contact with Basque' and 'attitudes towards Basque and the Basque ethnolinguistic group'. The correlation was that the higher the degree of contact with Basque and the more positive the attitudes towards Basque and the Basque ethnolinguistic group, the greater the frequency of occurrence of final falling intonational contours in information-seeking absolute interrogatives. The interpretation of this correlation was that the adoption of the characteristic Basque prosody allows speakers to be recognized as members of the Basque community. In the present study, we focused on rural areas. Falling intonational contours at the end of information-seeking absolute interrogatives were even more common than in urban areas (93.4%), and no correlation was found with degree of contact with Basque and with attitudes towards Basque. Our interpretation is that in rural areas the presence of Basque in daily life is stronger, and that there is a consolidated variety of Spanish used by all speakers regardless of their attitudes. Thus, the adoption of intonating features of this language is not the only indicator belonging to the Basque ethnolinguistic group. Our study reveals the great relevance of subjective social factors, such as language attitudes, in the degree of convergence between two languages.

Keywords: intonation; language contact; bilingualism; language attitudes; social factors; Spanish; Basque

1. Introduction

Within the growing field of research on phonetic and phonological issues of language contact and bilingualism, aspects of suprasegmental phonology have started receiving more attention, especially in prosody and intonation (a comprehensive list of references is provided in Elordieta and Romera (2020a)). A particularly interesting issue is that the presence of features of a language variety (LV-A) in another language variety (LV-B) is variable within the contact population. The main goal of this paper is to show that individual social factors of speakers of LV-A may help explain the differences among speakers in the degree of presence of LV-B features. Such factors can be the degree of contact of LV-A speakers with LV-B speakers and the attitudes of LV-A speakers towards LV-B and the LV-B ethnolinguistic group.

Sociolinguistic studies commonly explain linguistic variation in contact situations as the result of two types of factors: internal tendencies of the languages that favor a linguistic change, and the external influence that one language exerts on another in contact situations (Winford 2005, 2014; Poplack and Levey 2010, among others). Aspects such as the extent of contact, the density of speakers of each language, the relative prestige of each language, or the knowledge and relative use of each language by the speakers are usually invoked to explain transfer of features or convergence. However, other factors of a psychosocial nature also play a determining role in the adoption of linguistic features by speakers of languages in contact. Work along these lines includes that of Romera and Elordieta (2013) on Catalan intonational features in Spanish, Kozminska (2019) on the presence of English intonational features on the variety of Polish spoken by Polish immigrants, and Elordieta and Romera (2020a) on falling intonational contours in yes/no questions typical of Basque in the variety of Spanish in the Basque Country.

Elordieta and Romera (2020a) investigated the presence of (rising–)falling intonational configurations at the end of information-seeking absolute interrogatives in Basque Spanish as a feature that is received from Basque, directly or indirectly. Our study focused on the cities of Bilbao and San Sebastian. These two cities have always been in contact with Basque, although Spanish is the dominant language. In San Sebastian, a vernacular variety of Gipuzkoan Basque is still spoken, and in Bilbao, a variety of northern Bizkaian was spoken until the beginning of the last century. Next to the two cities, there are towns where Bizkaian and Gipuzkoan Basque are still spoken, and historically, there has been close contact with inhabitants of those towns in the form of commercial relationships and local immigration to the two capital cities. Bilbao and San Sebastian have populations of 343,430 and 181,652, respectively (858,236 and 327,428 when the urban areas they form together with smaller towns that surround them are considered; cf. (Eustat 2019)). The 1,185,664 inhabitants of the Bilbao and San Sebastian metropolitan areas amount to 54% of the population of the Autonomous Community of the Basque Country (2,188,017).

Elordieta and Romera (2020a) found that 79% of all information-seeking yes/no questions had final falling configurations in these two cities. The circumflex nuclear contour of information-seeking yes/no interrogatives in Basque Spanish is found in Castilian Spanish in absolute interrogatives that are not information seeking in nature, but rather have other pragmatic nuances, at least in elicited speech (from read sentences or from Discourse Completion Tasks). Escandell Vidal (1998, 1999, 2017) claims that a circumflex contour can appear in yes/no echo questions, used to express surprise at what the interlocutor has just said or to beg a clarification from the interlocutor of what (s)he has just uttered (cf. also (Estebas-Vilaplana and Prieto 2010; Hualde and Prieto 2015, p. 378)). Torreira and Floyd (2012) find this contour in yes/no questions that serve the discourse function of signaling that the topic of the discourse is being followed up, or that the “course of action” is maintained. In our corpus, the absolute questions were of the genuine information-seeking type, which, in Castilian Spanish, have been reported to have a final rising configuration. In order to be able to establish a more direct comparison with central varieties of Peninsular Spanish, we recorded seven speakers from Madrid in conversations of the same type as those in the Basque Country. The results showed that two-thirds of the information-seeking absolute interrogatives ended in a rising configuration (a percentage that rose to 84% for five of the seven speakers). Thus, the relative frequency of appearance of rising and falling configurations was roughly the opposite in Basque Spanish and Madrid Spanish.

This influence could be understood in diachronic terms as a historical transfer by native Basque speakers to their Spanish, followed by a consolidation of falling contours as a characteristic of Basque Spanish. Other northern varieties of Spanish, such as those spoken in Galicia, Asturias, and Cantabria, present falling final contours in absolute interrogatives, especially in non-urban areas. However, the falling contours are different from those found in Basque Spanish. For a detailed comparison, the reader is referred to Elordieta and Romera (2020a).[1]

In order to explain this variation, we tested several social factors, namely the attitudes towards the Basque language and the Basque ethnolinguistic group and the degree of contact with Basque. Our results showed that the attitudes speakers presented provided a high degree of explanation for the prosodic convergence of Spanish and Basque in two cities that were analyzed (Elordieta and Romera 2020a). Speakers who showed more positive attitudes towards Basque and the Basque-speaking group also produced higher rates of prosodic features present in Basque. The degree of contact was also a determining factor in explaining convergence of prosodic features. Together, the degree of contact with Basque and the attitudes toward this language explained almost 80% of the variation.

The one question that arises now is whether the situation is different in smaller towns where a vernacular variety of Basque has always been dominant. First, it is worth investigating whether a higher degree of contact with Basque and the Basque ethnolinguistic group in non-urban towns determines an even higher use of final falls in yes/no questions. Second, if the population in non-urban towns has positive attitudes towards Basque and the Basque ethnolinguistic group, it would be interesting to know whether this fact also leads to higher percentages of occurrence of falling nuclear contours in absolute interrogatives. These questions are the goal of our present paper.

2. Previous Study on the Prosody of Spanish in Contact with Basque and the Influence of Social Factors

2.1. Final Falling Contours in Absolute Interrogatives in Bilbao and San Sebastian

In Elordieta and Romera (2020a), we conducted a study on the intonation of information-seeking absolute interrogatives in Spanish as spoken in Bilbao and San Sebastian. We collected data through sociolinguistic interviews (Silva-Corvalán 2001) from 12 speakers of different linguistic profiles: Spanish monolinguals, L1 Spanish/L2 Basque bilinguals, and L1 Basque/L2 Spanish bilinguals. There were six females and six males, all between 35 and 55 years old and with secondary education at least.

The study revealed that in Bilbao and San Sebastian, 79% of all information-seeking yes/no questions (136 of a total of 172) had final configurations with a rising–falling circumflex contour. This contour can be transcribed in the autosegmental–metrical annotation system as L+(¡)H* (H)L%, that is, a rising pitch accent with the peak on the stressed syllable followed by a drop in tone in the final syllable. The pitch reached in the stressed syllable may exceed the level reached in the rest of the sentence, hence the upstep diacritic '¡'. On the other hand, the high tonal level may be maintained in the final syllable of the interrogative sentence and may fall even more abruptly towards the end, hence the possible presence of the high tone H in the boundary tone (subject to intra- and inter-speaker variation). Figure 1 shows an intonation contour of an absolute interrogative sentence in Basque Spanish, corresponding to a male bilingual speaker from San Sebastian with Spanish as his native language.

[1] There is a growing literature on the presence of prosodic features of one language on another it is in contact with. For a comprehensive bibliography, which includes studies on Spanish in contact with other languages in the Iberian Peninsula and in America, the reader can consult Elordieta and Romera (2020a).

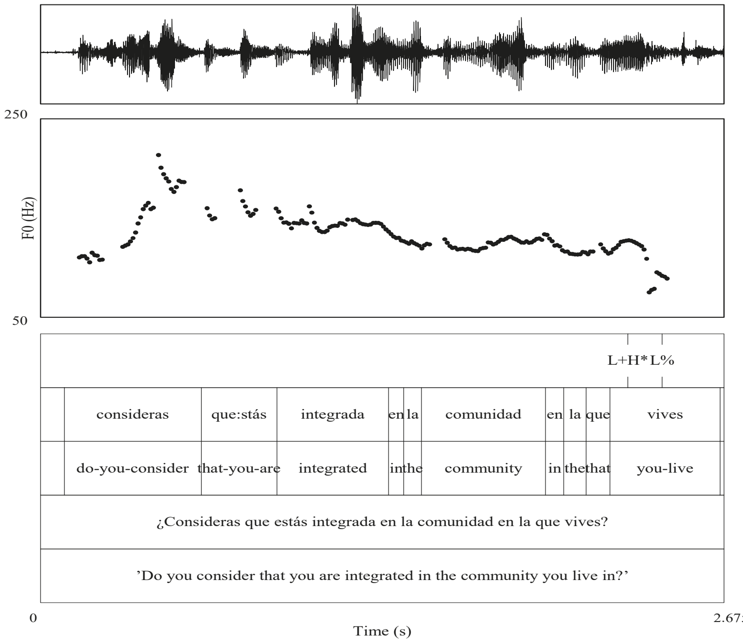


Figure 1. F0 contour of an absolute interrogative statement in Basque Spanish by a male L1 Spanish/L2 Basque speaker from San Sebastian.

Final rising or sustained pitch configurations are found in 21% of the information-seeking yes/no questions (36 interrogatives). Of these, only 10 interrogatives present the L* H% configuration of Castilian Spanish (6% of the total number of absolute questions), the rest having a rising nuclear accent, L+(¡)H* (¡)H%.

The finding that only 21% of information-seeking absolute interrogatives end in rising nuclear contours in Basque Spanish contrasts with the traditional description of neutral information-seeking absolute interrogatives in Castilian Spanish (central and southern varieties). They are characterized primarily by rising final contours in which the final stressed syllable of the statement presents a low tonal value followed by a rising intonation in the last syllable (see, among others, Navarro Tomás (1944), Quilis (1993), Face (2008), Hualde (2005), Estebas-Vilaplana and Prieto (2010), Henriksen (2010), Henriksen and García-Amaya (2012), and Hualde and Prieto (2015)). In Sp_ToBI, this tone is transcribed as L* H%.

Given that falling circumflex tones in absolute interrogatives are typical of Basque (cf. (Elordieta 2003; Gaminde et al. 2016; Robles-Puente 2012; Elordieta and Hualde 2014)), Elordieta and Romera (2020a) attributed the high frequency of falling circumflex tones in the Spanish spoken in the Basque Country to an influence from Basque (cf. also (Robles-Puente 2012; González and Reglero 2021)). Figure 2 shows an intonation contour of an absolute interrogative statement in Gipuzkoan Basque (example taken from (Elordieta and Hualde 2014, p. 457)).

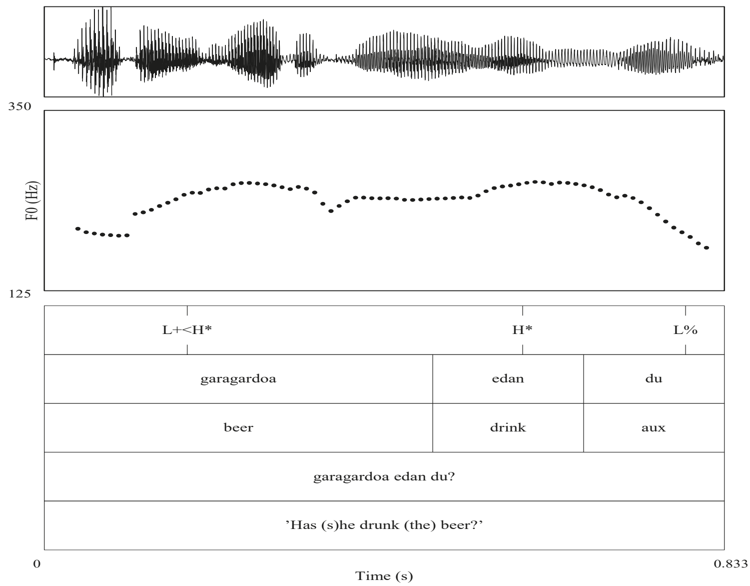


Figure 2. F0 contour of an absolute interrogative statement in Gipuzkoan Basque by a female speaker (example from (Elordieta and Hualde 2014, p. 457)).

Several studies have reported the existence of falling intonation patterns in absolute interrogatives in Central Castilian Spanish. These interrogative sentences are not genuine information-seeking questions, but rather have pragmatic connotations of echo, imperative, or confirmatory questions, in which the speaker is attributing the content of the interrogative to another person (cf. (Escandell Vidal 1998, 1999, 2017; Estebas-Vilaplana and Prieto 2010; Hualde and Prieto 2015; Henriksen et al. 2016), among others). These are annotated by the above-mentioned sources as L+H* L%, with a rise in pitch on the nuclear syllable above the level of other previous high tones (hence also annotated as L+¡H* L% by (Torreira and Floyd 2012)).

In spontaneous speech in Madrid Spanish, Torreira and Floyd (2012) claim that circumflex tones may be even more common than rising tones, which suggests that the most neutral intonation patterns are not necessarily the most common in conversational speech (cf. (Hualde and Prieto 2015)). Torreira and Floyd (2012) mention several but disperse pragmatic and discourse contexts where the circumflex contour occurs, but they come up with the generalization that this type of contour is mainly used as a "topic follow-up", and secondarily as a signal that the speaker is "maintaining the course of action". These interrogatives appear with the following functions in discourse in the corpus analyzed by these authors: responding to a previous question, providing receipt of news, initiating a repair, checking the listener's attention during a statement, or providing a pre-announcement during a statement.

Henriksen et al. (2016) also observed that rising tones are rare in spontaneous speech among speakers of Manchego Spanish, a variety of Castilian Spanish spoken to the south of Madrid. Inspired by Escandell Vidal (1998), the authors associated the rising contours with statements in which the content of the question is attributable to the speaker (in other words, a genuine information-seeking question). Falling contours are more common in interrogative sentences in which the content of the question can be attributed to another person, be it the speaker's conversation partner or another, external party.

The absolute interrogatives analyzed by Elordieta and Romera (2020a) in Basque Spanish, however, corresponded to a genuine search for information. The absolute interrogative statements occurred in the context of a semi-directed interview or conversation in which the interviewer asked questions that sought information about the interviewee that was unknown to the interviewer. Thus, the comparison remains legitimate: The unmarked pattern in Castilian Spanish for neutral, information-seeking absolute interrogatives is a final rise, but in Basque Spanish, it is a final fall. In any case, with the aim of settling the issue, Elordieta and Romera (2020b) carried out an analysis of the nuclear configurations of absolute interrogatives in Madrid Spanish by applying the same methodology followed in Elordieta and Romera (2020a) to seven speakers of Madrid Spanish so that the data could be directly comparable. The main result in Elordieta and Romera (2020b) was that in Madrid Spanish, 66.3% of information-seeking questions ended in rising contours, two-thirds of the total. In Basque Spanish, Elordieta and Romera (2020a) had found only 21% of information-seeking absolute interrogatives ending in a rising contour. That is, whereas final falls are the norm in Basque Spanish (79%), final rises are the norm in Madrid Spanish (only 33.7% of final falls). The pervasive presence of final falling configurations in Basque Spanish is such that there are no subjects with fewer than 64% of rising–falling nuclear contours, and two of the twelve subjects analyzed had all interrogatives ending in this type of contour, i.e., 100% of final falls, with no final rises at all.

2.2. The Role of Degree of Contact with Basque and Linguistic Attitudes towards Basque on the Prosody of Spanish in Contact with Basque in Urban Areas

In principle, one could consider the hypothesis that there could exist a correlation between the frequency of occurrence of final falling contours and the degree of knowledge of Basque. However, as shown in Table 1 below, Elordieta and Romera (2020a) found no significant differences in the frequency of occurrence of falling or rising contours in absolute interrogatives depending on the linguistic profile of the subjects (i.e., whether they were monolingual speakers of Spanish, L1 Spanish/L2 Basque speakers, or L1 Basque/L2 Spanish speakers). There were no differences depending on the city of origin (Bilbao or San Sebastian) or on the gender of the speakers, either.

Table 1. Numbers and percentages of information-seeking yes/no interrogatives with rising–falling and rising final contours (Falls and Rises, respectively) for each linguistic profile in San Sebastian and Bilbao.

	Monolingual Spanish		**L1 Spanish/L2 Basque**		**L1 Basque/L2 Spanish**	
	SSeb	**Bilbao**	**SSeb**	**Bilbao**	**SSeb**	**Bilbao**
Falls	19 (70%)	22 (73%)	22 (76%)	25 (89%)	28 (85%)	18 (78%)
Rises	8 (30%)	8 (27%)	7 (24%)	3 (11%)	5 (15%)	5 (22%)

These data indicate that at least 70% of interrogatives of all speakers ended in a fall in all groups. Although the monolingual group had a lower production of falling contours as compared to the other groups, no statistical correlation was found with the linguistic profiles (chi-square = 12.000; $p = 0.285$ for Bilbao; and chi square = 6.000; $p = 0.306$ for San Sebastian).

Rather, Elordieta and Romera (2020a) found a stronger relationship between the frequency of occurrence of final falling configurations with individual social factors, such as (a) the attitudes that each individual has towards Basque and the Basque ethnolinguistic group, and (b) the degree of contact of each individual with Basque and the Basque ethnolinguistic group. With respect to attitudes, all speakers expressed positive attitudes towards the Basque language and the Basque group in both cities. A total of 75.1% of the speakers showed very positive attitudes and only 24.9% showed not-so-positive attitudes. However, it was precisely the speakers who showed not-so-positive attitudes who also showed less production of downfalls (between 64% and 75% of falling interrogatives), as opposed to speakers more inclined towards Basque, whose production of falls was higher (77% to 100%). Despite this, the attitudinal factor could only account for roughly 50% of the variation ($R^2 = 0.466$, $F(1, 10) = 8.721$, $p = 0.01$). In contrast, the degree of contact accounted for 70% of the variation. Speakers with lower contact levels produced lower percentages of final falling contours (between 64% and 80%), while speakers with a higher contact value produced between 83% and 100% of final falls.

The degree of contact, however, seems to be a relevant factor in explaining the linguistic behavior of certain L1 Basque speakers, whose attitudinal value is positive but who had a percentage of interrogative falling circumflex contours relatively lower than expected for this attitudinal value. The substantial exposure not only to Basque Spanish, but also to other Peninsular Spanish varieties, as business trips to other parts of Spain might lead to producing more interrogative utterances with this type of final contour.

Finally, when both social factors were combined, 80% of the differences in the percentages of circumflex contours could be accounted for. That is, those speakers who had a closer contact with Basque or with speakers of Basque and who had favorable attitudes towards Basque and speakers of Basque presented higher percentages of final falling intonational contours (R^2 = 0.807, F (2, 9) = 18.844; p = 0.001). Therefore, although the degree of contact may account for a higher percentage of the production of falling contours, the combination of both factors provided a better explanation of the results.

3. Spanish Prosody and Social Factors in Non-Urban Basque-Speaking Areas

3.1. Methodology

For the present study, we recorded speakers from two rural or non-urban small towns: Lekeitio, in the same province as Bilbao, and Ibarra, in the same province as San Sebastian. Their populations as of 2020 were 7227 and 4306, respectively. The criteria for speaker selection were the same as those followed for Bilbao and San Sebastian. There were six speakers in each town (i.e., 12 in total), and they belonged to the same three linguistic profiles as in Bilbao and San Sebastian: Spanish monolinguals, L1 Spanish/L2 Basque bilinguals, and L1 Basque/L2 Spanish bilinguals. Gender, age, and level of education were also taken into account. Therefore, six women and six men between the ages of 35 and 55 years old and with a medium–high education level were recorded. (All subjects gave their informed consent for inclusion before they participated in the study. The ethic code for the research: CEISH/115/2012/ELORDIETA ALCIBAR)

The speakers were recorded using the same sociolinguistic interviews as for Bilbao and San Sebastian (Silva-Corvalán 2001). Two interviews were conducted with each experimental subject. The first interview was one in which the subjects were asked questions by the interviewers. With such an interview, declarative utterances were obtained from the subjects. The questions were divided into three modules. The first module contained questions on the subjects' degree of knowledge of the two languages in contact, Basque and Spanish. The second module had questions related to the degree of use of each of these languages. In the third module, the subjects were questioned on their attitudes towards Basque and the Basque ethnolinguistic group: for instance, whether they thought that speaking Basque improved the social image of a person, whether, in their opinion, knowing and speaking Basque was useful in their personal and professional lives, or whether they thought that Basque should be taught obligatorily. The questions were written on sheets of paper in the form of bulleted topics rather than in the form of full questions, the idea being that the interviewer posed the questions in as natural a style as possible, not as read speech.

The second interview was one in which the subjects took the role of interviewers and asked their interlocutors (i.e., the interviewers of the first part) the same questions that they had been asked. This way, a number of absolute and partial interrogative utterances (i.e., yes/no and wh-questions, respectively) were recorded from the subjects. The interviewers were speakers of Basque Spanish and members of the community, and had been trained to conduct the interviews. That way, we fostered a situation in which our experimental subjects could feel more at ease with the activity, talking to a person that speaks the same variety of Spanish as theirs. This could favor a more natural production from the subjects. The interviewers also served the role of selecting speakers who fulfilled the social and linguistic profiles described above.

The recordings took place in quiet rooms at the speakers' homes or workplaces in order to facilitate an optimal level of confidence and comfort for the speaker. The interviews were recorded with a Tascam DR-100 digital recorder through a built-in omni-directional microphone pointing towards the subjects but able to capture the speech of the two participants. The audio was recorded with a sampling rate of 44,000 Hz in wav format. The conversations were also recorded on video with a Sony video camera held on a tripod, with the objective of analyzing the speakers' level of (dis)comfort, relaxation, or nervousness while answering the questions about their attitudes towards Basque and the Basque ethnolinguistic group. In all, 9 h and 25 min of conversations were recorded for the 12 subjects, with an average of 47 min per subject. This is a remarkable similarity with our recordings in Bilbao and San Sebastian, where we collected 9 h and 20 min of conversations for 12 speakers, with an average of 46 min per speaker. A total of 360 declarative utterances, 155 absolute interrogatives, and 201 partial interrogatives were segmented—a similar amount in comparison with the one for Bilbao and San Sebastian (albeit with fewer absolute interrogative utterances). As for Bilbao and San Sebastian, we collected more declarative utterances than interrogatives because speakers produce more utterances when they respond than when they ask. We collected an average of 30 declarative utterances as well as almost 13 absolute interrogatives and 17 partial interrogatives.

Like for Bilbao and San Sebastian, in this article, only neutral or information-seeking absolute interrogatives were considered, without any pragmatic bias on the part of the speaker uttering the question. That is, the interrogative sentences corresponded to a genuine search for information. This choice was motivated by the fact that these are the sentence types that have the most significantly different intonational contours in Castilian Spanish and Basque. Hence, any intonational features in yes/no interrogatives in the Spanish variety of the Basque Country that differ from the typical Castilian Spanish features and which resemble those of Basque are more easily discernible. They commonly end in rising contours in the former (cf. (Face 2008; Estebas-Vilaplana and Prieto 2010; Hualde and Prieto 2015; Henriksen et al. 2016; Elordieta and Romera 2020b)) and in falling contours in the latter (cf. (Elordieta 2003; Gaminde et al. 2016; Robles-Puente 2012; Elordieta and Hualde 2014; Eguskiza et al. 2017)). A higher frequency of final falling intonational configurations would suggest an influence from Basque. In this regard, Elordieta and Romera's (2020b) study on information-seeking absolute interrogatives in Madrid Spanish is eloquent, as it uses the same methodology as the one in Elordieta and Romera (2020a), and the results are thus directly comparable. In this variety of Castilian Spanish, two-thirds of the information-seeking yes/no questions end in a rising contour, and thus only one-third of these types of questions end in a falling contour.

With respect to the social factors analyzed, the degree of contact, and the attitudinal component, all answers were coded both qualitative and quantitatively. From a quantitative point of view, each speaker was given two values, one regarding the degree of contact they maintained with the Basque language, and other regarding the attitude they showed towards the Basque language and the Basque ethnolinguistic group. Following studies in the field, each response was given a value on a scale of 1 to 3, 1 being the lowest value and 3 the highest. Then, the mean value of these ratings was calculated for each speaker, and the resulting number was assigned to each of them and taken as an index. We called these the contact value and attitudinal value, respectively (Elordieta and Romera 2020a, p. 25).

Finally, although the number of tokens analyzed may be judged to be low, it must be taken into account that conversational speech is not so straightforward to analyze. Overlaps, distortions, and interruptions in natural speech hinder the task of segmentation. The annotation of prosodic features is complicated as well due to uncontrolled phonological processes of resyllabification and phonetic lenition of consonants and vowels frequently observed in informal and fast speech. Moreover, the phonetic analysis was combined with a sociolinguistic analysis of the 12 speakers. In the end, the size of the data was roughly similar to the one for San Sebastian and Bilbao, and it will hence allow for a direct comparison as well.

3.2. Final Intonational Configurations in Absolute Interrogatives in Lekeitio and Ibarra

Of the 155 absolute questions in Lekeitio and Ibarra, we finally analyzed 137 utterances. We had to discard 18 interrogatives because they did not represent genuine information-seeking absolute interrogatives. Some of them were disjunctive (i.e., 'do you like X or do you prefer Y'?), and others were uttered without a verb, as not-full-fledged questions (e.g., 'place of origin?'). In the two towns, 93.4% of all information-seeking yes/no questions ended in final falling contours (128 out of 137). There was a slightly bigger percentage of falls in Lekeitio than in Ibarra, but the difference was small and, hence, not significant: 95.5% in Lekeitio (64 final falls out of 67 interrogatives) and 91.4% in Ibarra (64 final falls out of 70 interrogatives). Only 6.6% of all information-seeking yes/no questions ended in rising contours (9 out of 137), the most frequent contour being L+(¡)H* (¡)H%, that is, a rising nuclear accent with a peak on the nuclear syllable followed by a further rise on the final syllable, sometimes reaching a very high level, hence the upstepped diacritic. We only found one case of the contour reported traditionally as the most frequent one for Castilian Spanish, L* H%, that is, a low tone on the nuclear syllable followed by a high boundary tone.

Of all the information-seeking interrogatives, 53.3% were of the rising–falling circumflex type found in San Sebastian and Bilbao, that is, a rising nuclear tone with a peak on the tonic syllable (L+H*) followed by an L% or HL% boundary tone. The latter is distinguished from plain L% because the high tone level of the nuclear accent is maintained up to the middle of the final syllable of the word (also final in the utterance) before it falls to a low pitch level. In many of the instances, an upstepped pitch level was observed on the accentual H* tone, which can be transcribed in a Tones and Break Indices (ToBI) model as L+(¡)H*. Thus, the general shape for the most common nuclear contour in information-seeking interrogatives in Lekeitio and Ibarra is L+(¡)H* (H)L%. As we said, this is also the most common type of contour in San Sebastian and Bilbao (69.5%, cf. Elordieta and Romera 2020a). Figure 3 shows a pitch track of an information-seeking yes/no question with an L+H* L% contour uttered by a male Spanish monolingual speaker from Lekeitio.

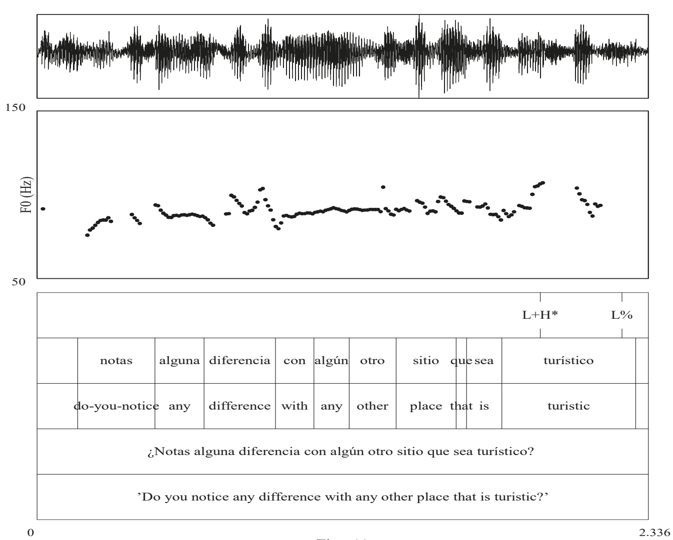


Figure 3. F0 contour of an absolute interrogative statement in Basque Spanish, by a male monolingual Spanish speaker from Lekeitio.

In several instances of final falling contours, the nuclear accent was not a rising one, or at least not clearly, and an L% boundary tone followed. We chose the label H* L% for this type of contour, shown by 38.7% of all the information-seeking absolute interrogatives. Figure 4 illustrates an example, uttered by a female L1 Basque speaker from Ibarra. The high F0 level at the beginning of the syllable with the nuclear accent (i.e., the syllable ke, in the word euskera 'Basque') is a microprosodic effect of the voiceless plosive [k].[2]

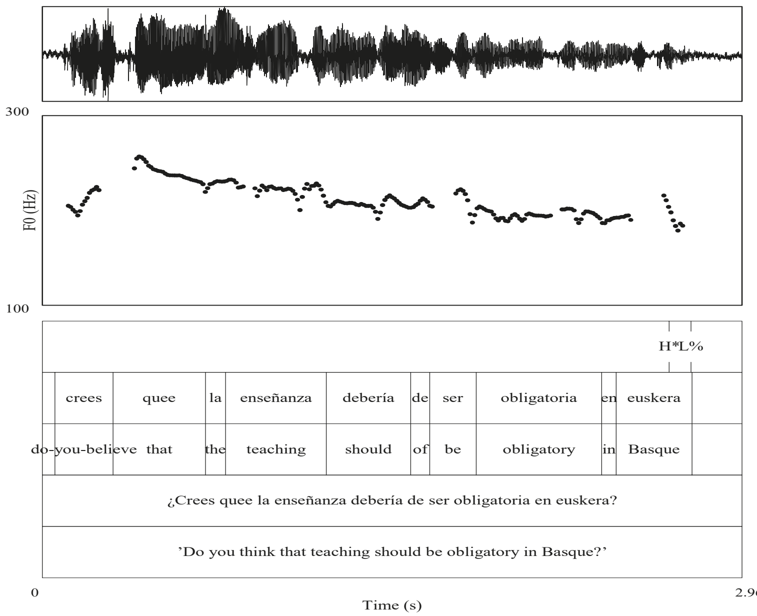


Figure 4. F0 contour of an absolute interrogative statement in Basque Spanish by a female L1 Basque/L2 Spanish speaker from Ibarra.

Finally, Figure 5 illustrates an F0 contour of an utterance with a final rising contour by a female L1 Spanish speaker from Lekeitio.[3]

---

2 The lengthening of word-final vowels is very common in spontaneous speech. In Figure 4, the final vowel of the complementizer *que* 'that' is written with two vowels to indicate that it is lengthened.

3 An anonymous reviewer asks whether the complexity of the syntactic structure of the interrogative utterances may influence the intonational contour. Since the utterance in Figure 5 (with a final rising configuration) is shorter and with simpler syntactic constituency than the ones in Figures 1, 3 and 4 (with a final falling configuration), the reviewer asks whether a correlation was found between the length and syntactic complexity of the utterances and the falling or rising final contours. However, there is no correlation. On the one hand, in our data there are many short utterances with final falls, such as *¿Estás casado?* 'Are you married?', *¿Tienes hijos?* 'Do you have children?', *¿Te gusta Lekeitio?* 'Do you like Lekeitio?', or *¿Te has fijado?* 'Have you noticed?'. On the other hand, among the very few utterances with final rises, there are long utterances such as *¿Y la persona que viva aquí, debería o tendría que hablar en euskera?* 'And the person that lives here, should (s)he or would (s)he have to speak in Basque?'.

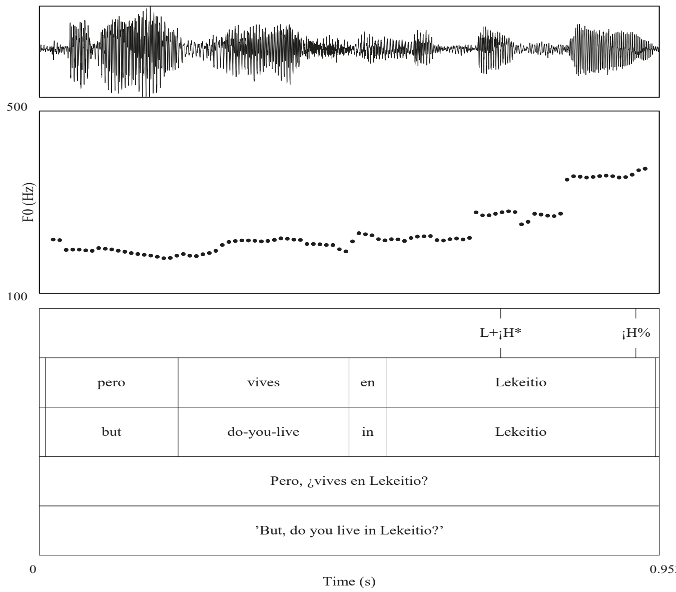


Figure 5. F0 contour of an absolute interrogative statement in Basque Spanish by a female L1 Spanish/L2 Basque speaker from Lekeitio.

3.3. Social Factors in Ibarra and Lekeitio

The fact that 93.4% of information-seeking absolute interrogatives in Ibarra and Lekeitio end in falling contours can be compared to the fact that 79% of yes/no questions of the same type end in falling contours in Bilbao and San Sebastian, as found by Elordieta and Romera (2020a). Within the Basque Country, then, it seems that falling configurations in information-seeking yes/no questions are more common in non-urban small towns than in cities. The difference is even bigger, of course, when compared to Madrid Spanish, where only 33.7% of information-seeking yes/no questions end in falls, as recently found by Elordieta and Romera (2020b). These results would suggest that the influence of Basque prosody is stronger in small towns than in cities.

Language dominance does not seem to have an effect on the occurrence of final falls or rises, like in Bilbao and San Sebastian (cf. Elordieta and Romera 2020a). Table 2 shows the percentages of final falling configurations depending on the speakers' linguistic profiles: monolingual Spanish, L1 Spanish, and L1 Basque. Although L1 Spanish speakers appear to have a smaller frequency of use of final falling configurations compared to the other two groups, the cross-tabulation test did not return a significant difference ($\chi^2 = 1.262$; $p = 0.532$). It is noteworthy that monolingual speakers produced more falling configurations than bilingual speakers.

Table 2. Numbers and percentages of information-seeking yes/no interrogatives with rising–falling and rising final contours (Falls and Rises, respectively) for each linguistic profile in the two towns combined: Lekeitio and Ibarra.

	Monolingual Spanish	L1 Spanish/L2 Basque	L1 Basque/L2 Spanish
Falls	49 (96%)	37 (90.2%)	42 (93.3%)
Rises	2 (4%)	4 (11.8%)	3 (6.7%)

Tables 3 and 4 show the frequencies of occurrence of final falls and rises in absolute interrogatives in Lekeitio and Ibarra, respectively. It is in Lekeitio where L1 Spanish bilinguals have fewer falls, but the differences with the other groups are not statistically significant ($\chi^2 = 2.844$; $p = 0.241$). In Ibarra, there is no difference between groups. It is rather telling that monolingual Spanish speakers in Lekeitio have 100% of final falling configurations, even more than L1 Basque speakers.

Like in Bilbao and San Sebastian, gender is not a significant factor. Males produced 71 information-seeking absolute interrogatives, of which 67 ended in a falling configuration and four ended in a rising contour (94.3% and 5.7%, respectively). Females produced 61 final falls and five final rises (92.4% and 7.6%, respectively). This similarity across genders holds in each town. Table 5 below shows the percentages of falling and rising contours according to the speakers' gender.

Table 3. Numbers and percentages of information-seeking yes/no interrogatives with rising–falling and rising final contours (Falls and Rises, respectively) for each linguistic profile in the town of Lekeitio.

	Monolingual Spanish	**L1 Spanish/L2 Basque**	**L1 Basque/L2 Spanish**
Falls	26 (100%)	17 (89.5%)	21 (95.4%)
Rises	0 (0%)	2 (10.5%)	1 (4.6%)

Table 4. Numbers and percentages of information-seeking yes/no interrogatives with rising–falling and rising final contours (Falls and Rises, respectively) for each linguistic profile in the town of Ibarra.

	Monolingual Spanish	**L1 Spanish/L2 Basque**	**L1 Basque/L2 Spanish**
Falls	23 (92%)	20 (91%)	21 (91.3%)
Rises	2 (8%)	2 (9%)	2 (8.7%)

Table 5. Percentage of information-seeking yes/no interrogatives with falling and rising final contours according to speakers' gender.

	Female	**Male**
Falls	61 (94.3%)	67 (92.4%)
Rises	5 (5.7%)	4 (7.6%)
Total	66 (100%)	71 (100%)

Consequently, our data indicate that in Lekeitio and in Ibarra, the production of falling interrogatives is the norm among the speakers interviewed. Table 6 below illustrates that for 75% of the speakers (9 out of 12), 90–100% of their information-seeking absolute interrogatives had final falling contours (these speakers are highlighted in gray). For 16% of the speakers (2 out of 12), 86–87.5% of this type of interrogative showed falling contours, and only one person had 78% of final falling configurations. That is, the frequency of occurrence of final falling intonational contours is 78–100%, with three-fourths of the speakers having frequencies of 90–100%. In fact, half the speakers in Ibarra and Lekeitio (6 out of 12, cf. Table 6) had all their information-seeking absolute interrogatives displaying final falling tonal configurations. These data indicate that the falling pattern is widespread in the area and that differences in the production of interrogatives are minimal.

In spite of this, we were interested in contrasting these results with those in the cities and seeing to what extent social factors were playing similar roles, i.e., whether the degree of contact and attitude towards the Basque language and the Basque ethnolinguistic group could still be playing a role in this minimal variation. The presence of Spanish in the cities of Bilbao and San Sebastian is stronger than in smaller, non-urban towns, where a vernacular variety of Basque has always been dominant. Hence, we would like to know whether a higher degree of contact with Basque and the Basque ethnolinguistic group in non-urban towns leads to a higher use of final falls in yes/no questions. In the same line, regarding the attitudinal factor, which proved relevant in explaining the variation in Bilbao and San Sebastian, we would like to know whether a positive attitude also led to higher percentages of occurrence of falling nuclear contours in absolute interrogatives in Lekeitio and Ibarra.

Table 6. Percentage of interrogative sentences ending in falling contours according to the speakers' language profile[4].

		Language Profile	n	%
% Falls	78.0	L1Sp	1	8.3%
	86.0	L1Bas	1	8.3%
	87.5	Mon	1	8.3%
	90.0	L1BasL1Sp	2	16.7%
	91.0	L1Sp	1	8.3%
	100.0	2Mon 2L1BasL1Sp	6	50%
Total			12	100%

With regard to the degree of contact with Basque, it showed correlation with the linguistic profile of the speakers ($p = 0.008$) (cf. Table 7, where the degree of contact is named "contact value"). On a scale from 1 to 3 (1 being the lowest degree of contact, and 3 the highest), monolingual speakers ranged between 1.22 and 1.44, while bilingual speakers (L1Sp-L2Bas; L1Bas-L2Sp) scored between 1.67 and 2.33. Monolingual speakers acknowledged having more limited contact with Basque, since although they partially understood the language, and their in-laws, some friends, and people at work could occasionally address them in Basque, their interactions took place only in Spanish. All of them had tried to study Basque at some point in their adult lives. Official schooling in Basque started in the 1990s, so given that the average age of these speakers was 45–55, they only received formal education in Spanish. All of them emphasized the effort and difficulty that studying Basque as an adult meant for them.

Table 7. Degree of contact according to the speakers' language profile and gender.

		Language Profile			**Total**
		Monolingual	**L_1Sp/L_2Bas**	**L_1Bas/ L_2Sp**	
Contact Value	1.22	2	1		3 (24.9%)
	1.33	1			1 (8.3%)
	1.44	1			1 (8.3%)
	1.67		1		1 (8.3%)
	1.89		1	1	2 (16.7%)
	2.22			3	3 (24.9%)
	2.33		1		1 (8.3%)
Total		4	4	4	12

However, as shown in Table 8 below, unlike in the urban areas, the degree of contact was not a relevant factor in explaining the differences in the production of falling interrogatives ($p = 0.1$). Three monolingual speakers and one L1Sp speaker whose degree of contact ranged between 1.22 and 1.44 produced all interrogatives (100%) in a falling contour. In contrast, the one L1Sp speaker with the highest degree of contact with Basque (2.33) performed only 78% of her interrogatives as falling.

4 In Tables 6–10, language profiles are abbreviated as follows: Mon for Monolingual, L1Sp for L1 Spanish, and L1Bas for L1 Basque.

Table 8. Percentage of interrogative sentences ending in falling contours according to degree of contact and speakers' language profile.

		Contact Value							Total
		1.22	**1.33**	**1.44**	**1.67**	**1.89**	**2.22**	**2.33**	
% Falls	78.0							L1Sp	1
	86.0				L1Bas				1
	87.5		Mon						1
	90.0				L1Sp		L1Bas		2
	91.0					L1Sp			1
	100.0	2 Mon L1Sp		Mon			2 L1Bas		6
Total		3	1	1	1	2	3	1	12

With respect to the other relevant social factor in urban areas, the attitudinal component, certain differences were found between monolingual and bilingual speakers (L1Sp and L1Bas). As shown in Table 9, the former presented slightly less favorable attitudes towards Basque and the Basque ethnolinguistic group (1.67–2.00), while the latter showed very favorable attitudes (2.33–2.56). Monolingual speakers considered that Basque was not very useful for them professionally or in their daily life, and that languages such as English should be promoted as much as Basque. The bilingual groups, on the other hand, considered Basque useful professionally and personally, and supported it being a language spoken by all inhabitants in the Basque Country. These differences, nonetheless, were not statistically significant (p = 0.200).

Table 9. Attitudinal values according to the speakers' language profile.

		Language Profile			Total
		Monolingual	**L_1Sp/L_2Bas**	**L_1Bas/L_2Sp**	
Attitude Value	1.67	1			1
	1.78	2	1		3
	2.00	1			1
	2.33		1	2	3
	2.44		1	2	3
	2.56		1		1
Total		4	4	4	12

However, no relation was found between the attitudes shown by the speakers and their production of falling contours in information-seeking absolute interrogatives (p = 0.300). Unlike in urban areas, where attitudes were clearly influential in the production of falling patterns (cf. Elordieta and Romera 2020a), this factor does not provide an explanation for the small differences in non-urban areas. That is, higher attitudinal values did not lead to higher occurrences of final falling contours. Three monolingual speakers with attitudinal values below 2.00 had 100% of information-seeking yes/no questions with a final descending contour, while bilingual speakers with attitudes above 2.00 showed lower percentages (cf. Table 10 below). It seems, therefore, that the production of falling intonational patterns in non-urban areas is conditioned by factors other than those governing urban areas.

Table 10. Percentage of interrogative sentences ending in falling contours according to attitudinal value and speakers' language profile.

		Attitudinal Value						Total
		1.67	1.78	2.00	2.33	2.44	2.56	
% Falls	78.0				L1Sp			1
	86.0				L1Bas			1
	87.5			Mon				1
	90.0				L1Bas	L1Sp		2
	91.0						L1Sp	1
	100.0	Mon	2 Mon L1Sp			2 L1Bas		6
Total		1	3	1	3	3	1	12

4. Two Factors: Type of Variety and Language as Index of Identity in Urban vs. Non-Urban Areas

As mentioned before, Elordieta and Romera (2020a) showed that two social factors were relevant for explaining the variation among speakers in frequency of occurrence of final falling intonational configurations in information-seeking absolute interrogatives in urban areas (Bilbao and San Sebastian). These factors were the degree of contact with Basque and the attitudes towards this language and the Basque ethnolinguistic group. That study found that a close degree of contact with Basque and a positive attitude towards this language and the Basque ethnolinguistic group were associated with a higher percentage of occurrence of falling nuclear contours, which are typical of Basque but not typical of Castilian Spanish.

In the present study, however, no correlation between the production of descending nuclear contours and social factors was found in non-urban areas (in the small towns of Ibarra and Lekeitio). We suggest that the explanation can lie in two fundamental aspects: first, the lack of real variation among speakers in the production of falling nuclear contours, and second, as a consequence of this, the impossibility of it representing any of the ethnolinguistic groups in these areas.

Regarding the first aspect, in the non-urban towns of Lekeitio and Ibarra, there are very small differences among speakers in the production of the intonational patterns of information-seeking absolute interrogative utterances. As shown in Elordieta and Romera (2020a), in the cities of San Sebastian and Bilbao, there were more differences among speakers (there were speakers with a 63% occurrence of falling intonational patterns, and others could have up to 100%). However, in the case of Ibarra and Lekeitio, the variation was minimal. The production of falling interrogatives ranged between 85% and 100% (except for one speaker, who had 79%), and half of them had 100%. Figure 6 below shows these differences between urban and non-urban populations.

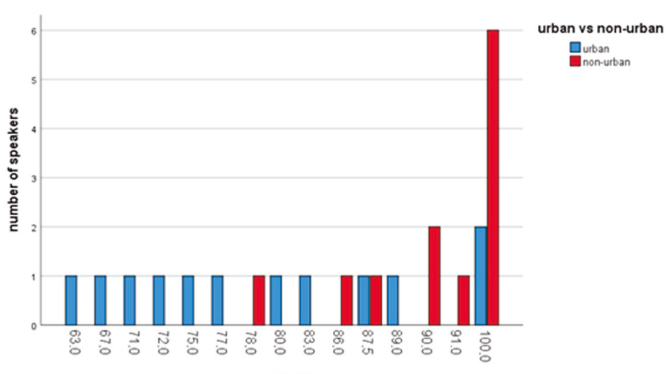


Figure 6. Percentage of interrogative sentences ending in a falling contour in urban and non-urban areas.

This lack of variation is characteristic of a process of settlement of a linguistic change (Blas-Arroyo and Lahoz 2018). Some synchronic social factors may help us establish that falling intonational contours at the end of information-seeking absolute questions are a feature of Basque Spanish in non-urban areas, where Basque has historically been the dominant language of use.

First, the generational group that was interviewed (35–55 years old) is the one that suffers the greatest social pressure to accommodate to a standard variety (Labov 2001). Information-seeking absolute interrogatives in central, standard Castilian Spanish are described to end in rising nuclear configurations, as already stated in the introductory section and in Section 2.1 of the present paper (cf. the references there). Our speakers have the opposite pattern, that is, (rising–)falling intonational contours at the end of such types of utterances. It thus seems that this feature is very stable in their variety. One could test the production of other generational groups, but as we know, younger and older generations (i.e., those between the ages of 18 and 30 and those above 60 years of age) do not have such a marked impact in the evolution of language change (Labov 2001). We therefore doubt that these groups can provoke substantial effects on the evolution of this feature.

Second, no gender differences were found in the production of falling yes/no questions. This reinforces the idea that such a feature is a settled one that does not respond to social variation.

A third argument that supports the view that falling intonational contours in absolute interrogatives are a stable feature of the variety of Spanish in non-urban towns is the historical and ongoing dominant presence of Basque. In the cities of Bilbao and San Sebastian, there is a much higher percentage of people for whom Spanish is the dominant language than in Ibarra and Lekeitio. In social terms, this means that in Bilbao and San Sebastian, the presence of the Spanish ethnolinguistic group is bigger than in Basque-dominant towns, such as Lekeitio and Ibarra. In these towns, there is little variation in the type of Spanish spoken. The variety there corresponds mainly to the vernacular variety spoken by L1 Basque speakers, the majority group in the area. The data provided by Eustat (2018) about the language used at home in 2016 reflect the overwhelming use of Spanish in the urban areas, as opposed to the dominance of Basque in the non-urban locations. As can be seen in Table 11 below, in Bilbao and in San Sebastian, the presence of Spanish at home is dominant, as opposed to Ibarra and Lekeitio, where the use of Basque at home amounts to almost 70%[5].

Table 11. Number of speakers who used Basque at home in 2016 (Eustat 2018).

	Basque	Spanish	Both	Other	Total
Bilbao	11,149	298,767	23,372	9784	343,072
San Sebastian	25,509	126,304	23,479	5366	180,658
Ibarra	1755	1633	618	129	4135
Lekeitio	4706	1758	565	177	7206

The variety of Spanish spoken in the Basque Country is the result of the long contact between these two languages (Elordieta and Romera 2020a). However, Spanish in urban areas is the result not only of the Spanish spoken by Basque speakers in the area, but also of the Spanish spoken by different groups who arrived in the Basque Country from other parts of Spain in the late 19th century and the 20th century (Zallo and Ayuso 2009). Rural areas, on the other hand, received less population, and Spanish speakers never outnumbered Basque speakers. Therefore, the Spanish spoken in non-urban areas is mostly the result of the variety of Spanish spoken by L1 Basque speakers, with a strong presence of Basque features. As we pointed out in Romera and Elordieta (2013), the transfer of features of one

5 The data reported here correspond to the use of Basque at home. We are aware that these data might differ from the use of Basque in other contexts. Unfortunately, only the global knowledge of Basque for populations of more than 40,000 inhabitants is available at this point (Eustat 2018), and although these data reduce the differences between the two languages, Spanish is still dominant in urban areas (Bilbao 61.18% and San Sebastian 57% for the population over 16).

language to another can occur not only by direct influence, but also as an indirect process. When L1 Basque speakers communicated with the Spanish-speaking populations arriving in the area, they used a variety of Spanish with presence of features of their L1 (Basque). In order to communicate, the Spanish speakers accommodated their variety to the Spanish of the L1 Basque speakers as well, and as a result, L1 Basque Spanish became the common variety spoken by all. Consequently, linguistic features of Basque will be more consolidated in the Spanish spoken in non-urban areas due to the lesser presence of Spanish speakers.

This process is responsible for the absence of variation in the prosodic pattern of interrogatives in Lekeitio and Ibarra. The falling interrogative pattern is an almost entirely consolidated feature, which no longer differentiates ethnolinguistic groups in the area. All speakers in these towns use it indistinctly, and therefore, it is no longer a marker of a social group. Rather, it is an indexical feature of geographical location (Chambers and Trudgill 1998). This explains the absence of correlation with social factors, such as linguistic profile, gender, degree of contact, or linguistic attitudes, since the feature has become characteristic of the Spanish variety of the area. In urban areas, on the other hand, the nuclear configurations of absolute interrogatives show more variation (cf. Table 10 above), and there is a correlation with social factors. The feature is not only a geographical index, but also a marker that attaches the speaker to either the Spanish or the Basque ethnolinguistic group. Since linguistic features are capable of being perceived as representative of a social group, they also can be used as an expression of the ethnolinguistic identity of the person who uses them.

The concept of identity has been widely debated through different theoretical frameworks. However, there is more or less agreement in distinguishing an individual identity and a social or collective identity. The former refers to the individual consideration of oneself, whereas the latter entails the definition of the subject as belonging to a group (Bucholtz and Hall 2005; Spencer-Oatey 2007). We claim that social identity is involved in the explanation of our data. Following Butler (1990), Bucholtz and Hall (2005), and Spencer-Oatey (2007), among others, we consider that social identity is expressed, constructed, and shown to others through concrete ways of speaking and acting.

The identities put into play are adapted to the relationships, and the features that the individuals project of themselves are emphasized or minimized according to the context. Linguistic forms are a primary way of expressing identity. As Tejerina (1999), Echeverria (2003), and Baxok et al. (2006) point out, language is the main element on which Basque identity is built in the Basque Country. In urban areas, where the group of Basque speakers is smaller and there is a large group of Spanish speakers, the expression of linguistic features associated with Basque has a clear indexicalizing value that serves to claim Basque identity. The choice of such features has a clear differentiating function, and it is made precisely by those speakers whose attitudes towards Basque and the Basque ethnolinguistic group are more positive (Elordieta and Romera 2020a). Those who use it clearly align themselves with the Basque language and culture, and adopt it in order to be identified as Basque.

In contrast, non-urban areas share a variety of Spanish that already presents a strong presence of Basque features, and all the speakers ascribe to it regardless of their linguistic profile. Therefore, the Spanish language ceases to differentiate one group from the other. Identity-based adscription in these areas cannot be approached in the same terms as in urban areas. Basque linguistic features in Spanish have no possibility of being used as an indexicalizing element of Basque identity, and therefore, the speakers' linguistic profile, the degree of contact with Basque, and the linguistic attitudes cannot have a correlation with them. All the speakers interviewed in these areas presented a strong Basque identity, and differences of identity are probably expressed at other levels (Spanish, European, etc.) (Azurmendi and Bourhis 1998). We leave this point to further investigation in the future.

In summary, we can say that the linguistic features associated with the Basque language are essential for claiming identity in urban areas, and therefore, we find a positive relationship between the use of these and positive attitudes. On the contrary, in non-urban areas, the linguistic features of Basque do not exert an indexicalizing function of Basque identity, since all speakers use them equally. This explains their lack of correlation with the social factors analyzed in this study.

5. Conclusions

In Elordieta and Romera (2020a), we found that an average of 79% of information-seeking absolute interrogatives ended in circumflex (rising–)falling intonational configurations in the varieties of Spanish spoken in the Basque cities of Bilbao and San Sebastian in northern Spain. In that study, a correlation was also revealed between the frequency of occurrence of final falling contours and two social factors: degree of contact with Basque and the attitudes towards this language and the Basque ethnolinguistic group. A close degree of contact with Basque and a positive attitude towards this language and the Basque ethnolinguistic group were associated with a higher percentage of occurrence of falling nuclear contours. The present study had the goal of continuing with our knowledge of the intonational characteristics of Basque Spanish by focusing on data from two non-urban areas in the Basque Country, the towns of Lekeitio and Ibarra. The research question was whether the diachronic and synchronic presence of vernacular varieties of Basque in these small towns may lead to different results in the frequency of occurrence of final falling intonational contours and the correlations with social factors.

Two main findings can be reported from the present study. On the one hand, a higher average frequency of final falling intonational contours at the end of information-seeking absolute interrogatives was found in the non-urban towns of Ibarra and Lekeitio compared to Bilbao and San Sebastian (93.4% vs. 79%). The same tonal configuration as in Bilbao and San Sebastian was observed: L+(¡)H* (H)L%. On the other hand, no correlation was observed between the production of descending nuclear contours and social factors. We suggested that the explanation of these differences between urban and non-urban areas lies in the lack of real variation among speakers in the production of falling nuclear contours. The variation is minimal, ranging from 85% to 100% final falling contours. Indeed, half of the speakers had 100% final falling configurations. The generational group that was interviewed (35–55 years old) should suffer the greatest social pressure to accommodate to a standard variety (Labov 2001), but these speakers hardly produced rising final contours as in Castilian Spanish. It is hence apparent that falling intonational contours at the end of information-seeking absolute questions are a solid and well-established feature of Basque Spanish in non-urban areas. We argue that the pervasive presence of a Basque intonational feature, such as a final fall in absolute interrogatives, must be due to the historical dominance of Basque as a language of use in non-urban areas. That is, L1 Basque Spanish (heavily influenced by Basque) became the common variety in non-urban areas. In urban areas, in contrast, Spanish is the dominant language. The variety of Spanish there shows influence from Basque, but other varieties of Castilian Spanish also came to be present with the arrival of immigrants from other parts of Spain.

In urban areas, the nuclear configurations of absolute interrogatives show more variation, and there is a correlation with social factors. Linguistic features are capable of being perceived as representative of a social group, and can be used as an expression of the ethnolinguistic identity of the person who uses them, understood in social or collective terms (cf. (Butler 1990; Bucholtz and Hall 2005; Spencer-Oatey 2007), among others). Indeed, Tejerina (1999), Echeverria (2003), and Baxok et al. (2006) hold that language is the main element on which Basque identity is built in the Basque Country. That is, apart from being an indexical feature of geographical location (Chambers and Trudgill 1998), using falling or rising final configurations in information-seeking absolute questions is also a marker that associates the speaker to either the Spanish or the Basque ethnolinguistic group.

In rural or non-urban areas, on the other hand, the prevalence almost exclusively of falling ends of information-seeking yes/no questions leads to the impossibility of identifying frequency of occurrences of such contours with any of the ethnolinguistic groups in these areas. All speakers in these towns use falling contours indistinctly, and therefore, final falls in these sentences are no longer a marker of a social group. This explains the absence of correlation with social factors, such as linguistic profile, gender, degree of contact, or linguistic attitudes.

Author Contributions: Conceptualization, M.R. and G.E.; methodology, M.R. and G.E.; validation, M.R. and G.E.; formal analysis, M.R. and G.E.; investigation M.R. and G.E.; resources, M.R. and G.E.; data curation, M.R. and G.E.; writing—original draft preparation, M.R. and G.E.; writing—review and editing, M.R. and G.E.; visualization, M.R. and G.E.; supervision, M.R. and G.E.; project administration, M.R. and G.E.; funding acquisition, M.R. and G.E. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been funded by the Ministry of Science and Innovation (grant number FFI2016-80021-P), the Basque Government (grant number IT1396-19) and the University of the Basque Country (grant number GIU18/221).

Acknowledgments: We are indebted to our speakers, without whom this work would not exist. We also wish to thank Varun DC Arrazola for help with the analysis of the data, and two anonymous reviewers for their positive and constructive comments. This paper also benefitted from valuable feedback provided by audiences at the following two conferences: *Monterey Bay Applied Linguistics Symposium 2019*, held at the University of California, Santa Cruz, 17 May 2019, and *I ALFALito "Dinámicas lingüísticas de las situaciones de contacto"*, held at the Universidad Autónoma de Madrid, 28–30 October 2019.

Conflicts of Interest: The authors declare no conflict of interest.

References

Azurmendi, María-José, and Richard Y. Bourhis. 1998. Presentation of the research project "ICYLCABE-1996": Cultural and linguistic identities in the bilingual Autonomous Communities (BAC) of Spain. *International Journal of Social Psychology* 13: 547–58. [CrossRef]

Blas-Arroyo, José Luis, and Francisco Javier Vellón Lahoz. 2018. On the trail of grammaticalization in progress: Has el que become a compound relative pronoun in the history of Spanish prepositional relative clauses? *Probus* 30: 1–45. [CrossRef]

Baxok, Erramun, Pantxoa Etxegoin, Terexa Lekunberri, Iñaki Martínez de Luna, Larraitz Mendizabal, Igor Ahedo, Xabier Itzaina, and Roldán Jimeno. 2006. *Euskal Nortasuna eta Kultura XXI. Mendearen Hasieran/Identidad y Cultura Vascas a Comienzos del Siglo XXI/Identité et Culture Basques au Début du XXIéme Siècle*. Bilbao: Eusko Ikaskuntza.

Bucholtz, Mary, and Kira Hall. 2005. Identity and interaction: A sociocultural linguistic approach. *Discourse Studies* 7: 585–614. [CrossRef]

Butler, Judith. 1990. *Gender Trouble. Feminism and the Subversion of Identity*. London: Routledge.

Chambers, Jack K., and Peter Trudgill. 1998. *Dialectology*. Cambridge: Cambridge University Press.

Echeverria, Begoña. 2003. Language ideologies and practices in (en)gendering the Basque nation. *Language in Society* 32: 383–413. [CrossRef]

Eguskiza, Naia, Aintzane Etxebarria, and Iñaki Gaminde. 2017. Bai-ez galderak. In *Bariazioa Esaldien Intonazioan*. Edited by Aintzane Etxebarria and Naia Eguskiza. Bilbao: Universidad del País Vasco/Euskal Herriko Unibertsitatea, pp. 93–127.

Elordieta, Gorka. 2003. Intonation. In *A Grammar of Basque*. Edited by José Ignacio Hualde and Jon Ortiz de Urbina. Berlin: Mouton de Gruyter, pp. 72–112.

Elordieta, Gorka, and Jose I. Hualde. 2014. Intonation in Basque. In *Prosodic Typology II: The Phonology of Intonation and Phrasing*. Edited by Sun-Ah Jun. Oxford and New York: Oxford University Press, pp. 405–63.

Elordieta, Gorka, and Magdalena Romera. 2020a. The influence of social factors on the prosody of Spanish in contact with Basque. *International Journal of Bilingualism*. OnlineFirst. [CrossRef]

Elordieta, Gorka, and Magdalena Romera. 2020b. The intonation of information-seeking absolute interrogatives in Madrid Spanish. *Estudios de Fonética Experimental* 29: 195–213.

Escandell Vidal, Victoria. 1998. Politeness: A relevant issue for relevance theory. *Revista Alicantina de Estudios Ingleses* 11: 45–57. [CrossRef]

Escandell Vidal, Victoria. 1999. Los enunciados interrogativos: Aspectos semánticos y pragmáticos. In *Gramática Descriptiva de la Lengua Española*. Edited by Ignacio Bosque and Violeta Demonte. Madrid: Espasa-Calpe, pp. 3929–91.

Escandell Vidal, Victoria. 2017. Intonation and evidentiality in Spanish polar interrogatives. *Language and Speech* 60: 224–41. [CrossRef] [PubMed]

Estebas-Vilaplana, Eva, and Pilar Prieto. 2010. Castilian Spanish intonation. In *Transcription of Intonation of the Spanish Language*. Edited by Pilar Prieto and Paolo Roseano. Munich: Lincom Europa, pp. 17–48.

Eustat (Euskal Estatistika Erakundea—Instituto Vasco de Estadística). 2018. Población de la C.A. de Euskadi por el Municipio de Residencia, Según el sexo y la Lengua Hablada en Casa. Available online: https://www.eustat.eus/elementos/ele0014700/poblacion-de-la-ca-de-euskadi-por-el-municipio-de-residencia-segun-el-sexo-y-la-lengua-hablada-en-casa/tbl0014755_c.html (accessed on 11 October 2020).

Eustat (Euskal Estatistika Erakundea—Instituto Vasco de Estadística). 2019. Datos Estadísticos de la Comunidad Autónoma de Euskadi. Available online: https://www.eustat.eus/elementos/ele0016700/la-poblacion-de-la-c-a-de-euskadi-alcanza-su-maximo-historico-a-1-de-enero-de-2019-con-2188/not0016799_c.html (accessed on 11 October 2020).

Face, Timothy L. 2008. *The Intonation of Castilian Spanish Declaratives and Absolute Interrogatives*. Munich: Lincom Europa.

Gaminde, Iñaki, Asier Romero Andonegi, Urtza Garay, and Aintzane Etxebarria. 2016. Los tonos de frontera de las oraciones interrogativas absolutas producidas por hablantes bilingües vasco-español. *Estudios de Lingüística Aplicada* 54: 61–79. [CrossRef]

González, Carolina, and Lara Reglero. 2021. The intonation of yes-no questions in Basque Spanish. In *LanguagePatterns in Spanish and Beyond*. Edited by Juan Colomina-Almiñana and Sandro Sessarego. London: Routledge. [CrossRef]

Henriksen, Nicholas, and Lorenzo García-Amaya. 2012. Transcription of intonation of Jerezano Andalusian Spanish. *Estudios de Fonética Experimental* 21: 109–62.

Henriksen, Nicholas, Meghan E. Armstrong, and Lorenzo García-Amaya. 2016. The intonational meaning of polar questions in Manchego Spanish spontaneous speech. In *Intonational Grammar in Ibero-Romance*. Edited by Meghan E. Armstrong, Nicholas Henriksen and Maria del Mar Vanrell. Amsterdam: John Benjamins Publishing Company, pp. 181–205.

Henriksen, Nicholas. 2010. Question Intonation in Manchego Peninsular Spanish. Ph.D. dissertation, Indiana University, Bloomington, Indiana.

Hualde, José Ignacio, and Pilar Prieto. 2015. Intonational variation in Spanish, European and American varieties. In *Intonation in Romance*. Edited by Sónia Frota and Pilar Prieto. Oxford: Oxford University Press, pp. 350–91.

Hualde, José Ignacio. 2005. *The Sounds of Spanish*. Cambridge: Cambridge University Press.

Kozminska, Kinga. 2019. Intonation, identity and contact-induced change among Polish-speaking migrants in the U.K. *Journal of Sociolinguistics* 23: 29–53. [CrossRef]

Labov, William. 2001. *Principles of Linguistic Change, Volume 2: Social Factors*. Hoboken: Wiley.

Navarro Tomás, Tomás. 1944. *Manual de Entonación Española*. Madrid: CSIC.

Poplack, Shana, and Stephen Levey. 2010. Contact-induced grammatical change: A cautionary tale. In *Language and Space: An International Handbook of Linguistic Variation*. Edited by Peter Auer and Jürgen Erich Schmidt. Berlin: Mouton de Gruyter, vol. I, pp. 391–419.

Quilis, Antonio. 1993. *Tratado de Fonología y Fonética Españolas*. Madrid: Gredos.

Robles-Puente, Sergio. 2012. Two languages, two intonations? Statements and yes/no questions in Spanish and Basque. *Anuario del Seminario de Filología Vasca Julio de Urquijo: International Journal of Basque Linguistics and Philology* 46: 251–62.

Romera, Magdalena, and Gorka Elordieta. 2013. Prosodic accommodation in language contact: Spanish intonation in Majorca. *International Journal of the Sociology of Language* 221: 127–51. [CrossRef]

Silva-Corvalán, Carmen. 2001. *Sociolingüística y Pragmática del Español*. Washington, DC: Georgetown University Press.

Spencer-Oatey, Helen. 2007. "Theories of identity and the analysis of face.". *Journal of Pragmatics* 39: 639–656. [CrossRef]

Tejerina, Benjamín. 1999. El poder de los símbolos: Identidad colectiva y movimiento etnolingüístico en el País Vasco. *Revista Española de Investigaciones Sociológicas* 88: 75–105. [CrossRef]

Torreira, Francisco, and Simeon Floyd. 2012. Intonational meaning: The case of Spanish yes–no questions. Paper presented at the Tone and Intonation in Europe, University of Oxford, Oxford, UK, September 6–8.

Winford, Donald. 2005. Contact-induced changes: Classification and processes. *Diachronica* 22: 373–427. [CrossRef]

Winford, Donald. 2014. Social factors in contact languages. In *Contact Languages*. Edited by Peter Bakker and Yaron Matras. Berlin: Mouton de Gruyter, pp. 363–416. [CrossRef]

Zallo, Ramón, and Mikel Ayuso. 2009. *Conocer el País Vasco: Viaje al Interior de su Cultura, Historia, Sociedad e Instituciones*. Bilbao: Servicio Central de Publicaciones del Gobierno Vasco.

languages

MDPI

Article

Asymmetry and Directionality in Catalan–Spanish Contact: Intervocalic Fricatives in Barcelona and Valencia

Justin Davidson

Department of Spanish and Portuguese, University of California, Berkeley, CA 94720, USA; justindavidson@berkeley.edu

Received: 2 September 2020; Accepted: 11 November 2020; Published: 13 November 2020

Abstract: Multilingual communities often exhibit asymmetry in directionality by which the majority language exerts greater influence on the minority language. In the case of Spanish in contact with Catalan, the asymmetry of directionality, favoring stronger influence of Spanish as a majority language over Catalan, is complicated by the unique sociolinguistic statuses afforded to different varieties of Catalan. In order to empirically substantiate the social underpinnings of directionality in language contact settings, the present study examines the variable voicing and devoicing of intervocalic alveolar fricatives in Spanish, Barcelonan Catalan, and Valencian Catalan as processes that are historically endogenous and equally linguistically motivated in both languages. Intervocalic fricatives in both languages were elicited using a phrase-list reading task, alongside sociolinguistic interviews for attitudinal data, administered to 96 Catalan–Spanish bilinguals stratified by gender, age, and language dominance in Barcelona and Valencia, Spain. Patterns of sociolinguistic stratification consistent with community-level changes in progress favoring either Catalan-like voicing or Spanish-like devoicing varied by community, with a stronger influence of Catalan on Spanish in Barcelona and Spanish on Catalan in Valencia. These asymmetries, corroborated by attitudinal differences afforded to Catalan and Spanish in Barcelona and Valencia, ultimately reinforce the role of social factors in language contact outcomes.

Keywords: multilingualism; agentivity; directionality; fricative (de)voicing; Catalan–Spanish contact; sociophonetics

---

1. Introduction

Observations of asymmetry and directionality with regard to language contact effects have long been addressed in linguistics research, applicable both at the level of the individual multilingual speaker, as well as at the broader level of the multilingual speech community. With regard to individual-level effects, crosslinguistic influence between a speaker's first language (henceforth L1) and second language (henceforth L2) is characterized by unequal (i.e., asymmetric) effects by which the L1 more strongly influences the L2 (i.e., directionality of L1 to L2) (Winford 2005, p. 373). Indeed, various phonological models of production and perception of L2 speech (cf. Best 1995; Best and Tyler 2007; Escudero 2005; Flege 1995) posit that L1 categories directly mediate the variable acquisition of L2 categories, which together attempt to account for the persistence of an L2 accent despite relatively early and even prolonged exposure and usage of the L2 (among many, Bosch et al. 2000; Pallier et al. 1997; Flege 2002; Flege et al. 1997, 1995, 2006; Flege and Munro 1994; Guion et al. 2000). At the level of the multilingual speech community, asymmetry and directionality have been characterized along a probabilistic hierarchy of contact influence whereby a majority language is likely to exert greater linguistic influence on a minority language as resultant from an array of typical social differences across L1-speaker groups and the languages themselves, such as population size (e.g., greater number of L1 speakers of the majority language), sociopolitical status (e.g., official status and linguistic capital afforded to majority language), sociocultural status (e.g., L1 speakers of the majority language as socioeconomically and culturally dominant), and language attitudes (e.g., more positive associations of power and linguistic vitality afforded to the majority language) (Thomason 2001, 2010; Thomason and Kaufman 1988).

Though the empirical investigation of crosslinguistic or contact influence has traditionally centered on cases of L1 to L2 directionality (or source language agentivity (Van Coetsem 2000)) or majority language to minority language directionality at the levels of the individual speaker and greater speech community, respectively, evidence of L2 to L1 directionality (or recipient language agentivity (Van Coetsem 2000)) and minority language to majority language directionality is robust. At the level of the individual speaker, for example, Flege (1987) found that French–English and English–French bilinguals developed a merged L1–L2 category with respect to the voiced onset time (henceforth VOT) of /t/, resulting, respectively, in a partially English-like L1-French /t/ and partially French-like L1-English /t/. Parallel cases regarding the VOT of English and Italian voiced stops by Italian–English and English–Italian bilinguals and the VOT of English stops (in addition to the first and second formant frequencies of select vowels) by L1-English L2-Korean bilinguals are respectively reported in MacKay et al. (2001) and Chang (2012), ultimately argued to evidence systematic phonetic interactions between the L1 and L2 categories in the shared phonetic sound space (Flege 2002). At the level of the speech community, L2 influence on an L1 is most predominantly documented with respect to lexical borrowing or the innovation of loanwords (Winford 2010). Cases of L2 influence in non-lexical domains, or structural borrowing (see, for example, Sanchez (2008)), have been posited to be either less common (Thomason and Kaufman 1988), highly constrained by the languages' grammars (Silva-Corvalán 1986), or perhaps altogether unattested (Poplack and Levey 2010).[1] Accordingly, to better address these asymmetries with regard to bidirectional (i.e., L1 to L2 and L2 to L1) contact effects, the present study explores a unique case of sociophonetic variation across bilingual speakers of Catalan and Spanish hailing from communities of distinct sociolinguistic status and language attitudes, operationalized with respect to measures of linguistic vitality and (c)overt associations of power and solidarity. This, alongside the selection of a phonetic variable equally motivated to appear in either language, permits an innovative analysis of the social underpinnings of community-level linguistic variation and change in multilingual communities.

2. Catalan and Spanish in Barcelona and Valencia

The sociopolitical histories between Spanish and Catalan involve centuries-old contact between the two languages, ultimately culminating in an 18th century shift from the previous state of societal monolingualism in Catalan (as a national language) to the declaration of Spanish as the sole language of the state, and indeed the compulsory acquisition of Spanish through public education in the 19th century (Vallverdú 1984, pp. 19–21; Vila-Pujol 2007, pp. 62–63). The rise of Spanish hegemony over Catalan reached a peak during Spain's fascist dictatorship under General Francisco Franco from 1939 until his death in 1975, during which legislation was actively passed to eliminate or otherwise Castilianize all non-Spanish institutions, as well as outlaw Catalan and other non-Spanish languages in the public sphere (Newman et al. 2008, p. 307; Turell Julià 2000, p. 47; Vallverdú 1984, p. 24; Vila-Pujol 2007, p. 64). The restoration of Catalan as a co-official language in the Autonomous Communities of Valencia, Catalonia, and the Balearic Islands came as a product of Spain's 1978 Democratic Constitution, shortly after which (in 1983) the Law of Linguistic Normalization and the Use and Teaching of Valencian

1 The polemic status of structural borrowing is rooted in competing viewpoints regarding language-internal (or endogenous) and language-external (or contact-induced) factors, which fall outside the scope of the present paper. See Thomason (2008) and references therein for a fuller discussion of these arguments.

Act (respectively for Catalonia and Valencia) restored Catalan as a vehicle for public education (Huguet 2006, p. 150; Newman et al. 2008, pp. 306–7; Vann 1999, pp. 317–18).

Despite the restoration of Catalan as a language of (co-)official status, the sociolinguistic trajectories of Catalan in Catalonia and Valencia have shown considerable degrees of divergence. In Catalonia, thanks in part to ample efforts on behalf of the local government and media to consistently promote Catalan's strong expansion throughout the public and legislative sectors (Pradilla 2001, pp. 63–65), Catalan is readily characterized as the language of local political and economic power, with Spanish being associated with the lower socioeconomic class and immigrant communities (Siguan 1988, p. 454; Sinner 2002, p. 161). A longitudinal series of language attitude research featuring the matched guise technique (Woolard 1984, 1989, 2009, 2011; Woolard and Gahng 1990; Newman et al. 2008) since the 1980s has shown that positive associations of the Catalan language, and even a Catalanized accent in Spanish, commonly index a bilingual, expressly Catalonian identity, tied overtly and covertly to attributes of solidarity in the community (Davidson 2019). Barcelona (city) 2011 census data show that self-reported competence in Catalan for understanding, reading, speaking, and writing are respectively 95%, 79%, 72%, and 53% (Institut d'Estadística de Catalunya 2014), which have steadily increased since the 1980s and reflect the considerable degree of linguistic vitality of this minority language (Pradilla 2001, p. 62).

The status of Valencian Catalan, on the other hand, contrasts rather directly with that of Barcelonan Catalan. Since 1995, efforts to restore the administrative and ideological status of Catalan to match (or even surpass) that of Spanish in Valencia have been actively curtailed by a series of conservative political party leaders who have aligned themselves with a group of pro-Spanish, Valencian elites that gained considerable power and wealth during the Franco regime (Casesnoves Ferrer 2010, pp. 479–80; Casesnoves Ferrer and Sankoff 2004, p. 2; Pradilla 2001, pp. 68–69). A highly successful propaganda campaign was launched against (Catalonian) Catalan, based off the fear that the growing Catalonian independence movement would subsume the Valencian state. Beyond disparaging ties to Catalonia and its speakers, this campaign additionally positioned Valencian as a completely unrelated language from (Catalonian) Catalan, which served to fuel a pro-Valencian (and specifically anti-Catalan) movement that was ideologically aligned with Spanish and the nation-state as symbols of anti-Catalan-ness, rather than Valencian (Casesnoves Ferrer 2010, p. 480; Pradilla 2001, pp. 69–70).[2] Under this campaign, Valencian has rarely been used in administrative contexts, and the once-thriving *Canal 9* Valencian TV station was shut down in 2012 (Pradilla 2001, p. 69). Matched guise research in Valencia has found that whereas positive, local affiliations of solidarity were originally (in 1998) afforded to Valencian, in 2008, these were newly afforded to Spanish in the capital city of Valencia (Casesnoves Ferrer 2010, p. 486). The 2011 census data for the aforementioned self-reported competences for Valencian Catalan in understanding, reading, speaking, and writing in the city of Valencia are respectively 89%, 61%, 48%, and 61% (Generalitat Valenciana 2011), which, when compared to the corresponding aforementioned census data for Barcelona, notably lag behind the most in terms of speaking competence.

Accordingly, Barcelona and Valencia present two unique sociolinguistic and sociopolitical realities for the same language contact pairing between Catalan as a minority language and Spanish as a majority language. While linguistic differences between these contact settings are not unilaterally determined from their distinct social contexts, their comparison nonetheless facilitates an empirical assessment of the contributions of these social differences to linguistic outcomes as concerns the notions of directionality and asymmetry of contact influence.

[2] Anecdotally, the visual landscapes of modern Valencia and Barcelona are quite telling. From my travels in 2018, the hanging of a *senyera* (the Catalan flag of nationhood and independence) off one's balcony has become an extremely prevalent practice in Barcelona. In Valencia, the analogous Valencian flag can only rarely be found, hidden amidst a sea of (national) Spanish flags adorning the balconies of the city's thoroughfares.

3. Alveolar Fricatives in Spanish and Catalan

North-Central Peninsular Spanish features an apical-alveolar voiceless /s/, articulated with a gesture of the tongue-tip toward the alveolar ridge (Hualde 2014, p. 147; Martínez Celdrán and Planas 2007, p. 110; Quilis 1981, pp. 234–35). In monolingual Spanish varieties that do not exhibit aspiration or deletion of /s/ in pre-consonantal positions, such as North-Central Peninsular Spanish (e.g., Barcelonan Spanish and Valencian Spanish), two allophones of /s/, namely voiceless [s] and voiced [z], are prescriptively found in complementary distribution via regressive assimilation of voicing to the following consonantal segment. Before voiced (semi)consonants, /s/ is realized as [z] (e.g., *rasgo* [ráz.ɣo] 'feature'; *mis hierbas* [miz.ɟ̞éɾ.βas] 'my herbs'), whereas in all other contexts, /s/ is produced as [s] (e.g., *rasco* [ráz.ko] 'I scratch'; *casa* [ká.sa] 'house'; *monos* [mó.nos] 'monkeys') (Hualde 2014, pp. 154–55; Morgan 2010, p. 248). Accordingly, monolingual Spanish productions of [z] outside of the context of a following voiced (semi)consonant (e.g., the intervocalic context in particular) are prescriptively disallowed:

> "*La s sonora aparece únicamente, en nuestra lengua, en posición final de sílaba, precediendo inmediatamente a otra consonante sonora; en cualquier otra posición su presencia es anormal y esporádica*" [The voiced /s/ in our language appears solely in syllable-final position immediately preceding another voiced consonant; in any other position, its presence is abnormal and sporadic]. (Navarro Tomás 1918, p. 83)

In contrast to Spanish, Catalan features two apical-alveolar fricative phonemes, voiceless /s/ and voiced /z/. This phonemic voicing contrast is active word-initially and word-medially, producing minimal pairs such as *zel* 'zeal' [zɛ́ɫ] / *cel* 'sky' [sɛ́ɫ] and *pesar* 'to weigh' [pə.zá] / *passar* 'to pass' [pə.sá]. Critically, this phonemic voicing contrast is neutralized word-finally, resulting in [s] or [z] depending on the voicing feature of the following segment (that is, the voicing neutralization of word-final Catalan alveolar fricatives (and, in fact, all Catalan sibilants) resolves by means of anticipatory assimilation). When followed by a voiced segment, such as a vowel, the word-final fricative is systematically voiced (e.g., *gos* [s] 'dog'; *gos estrany* [z] 'strange dog') (Hualde 1992, pp. 371–72, 393–94; Hualde and Prieto 2014, p. 109; Recasens 2014, pp. 239–40; Wheeler 2005, pp. 147–49, 162).

Accordingly, voiced intervocalic fricatives in Catalan are resultant from word-initial /z/, word-medial /z/, and as a product of voicing assimilation of word-final prevocalic /s/ and /z/ (or archiphoneme /S/)). This accordingly sets up an interesting pair of opportunities for bidirectional contact influence contingent on syllable position. With respect to syllable-initial contexts, productions of Spanish *pesar* 'to weigh' or *casa* 'house' as [pe.záɾ] and [ká.za] on the part of an L1-Catalan speaker could evidence the transfer of a Catalan phoneme (/z/) into Spanish, whereas productions of Catalan *pesar* 'to weigh' or *casa* 'house' as [pə.sá] and [ká.sə] on the part of an L1-Spanish speaker could evidence the substitution of Spanish /s/ for Catalan /z/, potentially eliminating the phonemic voicing contrast in Catalan. With respect to word-final contexts, the production of Spanish *las albas* 'the dawns' as [la.zál.βas] by an L1-Catalan speaker or the production of Catalan *les albes* 'the dawns' as [lə.sál.βəs] by an L1-Spanish speaker would constitute a case of largely phonetic, rather than phonemic, transfer (i.e., the respective transfer of a Catalan or Spanish phonotactic voicing rule, which would not create or eliminate any phonological contrasts).[3]

Notably, though the phonological voicing contrast between Catalan /s/ and /z/ is a feature of the prescriptive, standardized academy norms for both Barcelonan Catalan (Julià i Muné 2008, pp. 66–67) and Valencian Catalan (Real Acadèmia de Cultura Valenciana 2000; Acadèmia Valenciana de la Llengua

[3] Though Catalan /z/ is sometimes framed as a novel L2 category for L1-Spanish learners to acquire (Carrera-Sabaté et al. 2016, p. 48), the existence of Spanish [z] before voiced consonants suggests that, rather than a case of foreign category acquisition, the present study instead entails the acquisition of novel phonotactic structure, wherein [z] is to appear in non-Spanish contexts (e.g., syllable-initially (contrastive with /s/) and prevocalic word-finally (non-contrastive with /s/)).

2006, p. 29), select oral vernaculars of Barcelonan and Valencian Catalan have been characterized as having lost the voicing contrast in favor of exclusively voiceless intervocalic alveolar sibilants. In a sociophonetic investigation of *xava* Catalan, a Barcelonan sociolect originally associated with the L1-Spanish-speaking working class, Ballart (2013, p. 145) finds that /z/ is realized as [s] with a frequency of 15% by L1-Catalan speakers, in comparison to the 58% rate of [s] production exhibited by L1-Spanish speakers. For Valencian Catalan, the regional vernacular known as *apitxat* is similarly characterized as lacking voiced /z/ (Prieto 2004, p. 216; Moll 2006, p. 109), deemed *no recomanable* ("not recommendable") by the Valencian Academy of Language (Acadèmia Valenciana de la Llengua 2006, p. 29). Ultimately, since prescriptive academy norms do not accurately reflect real language use, the existence of *xava* and *apitxat* do not hinder the present investigation concerning intervocalic fricative production in Barcelona and Valencia, and instead are indicative of the pervasive reality of sociolinguistic variation at even the phonological level, which I aim to expressly link to select social and linguistic factors. Indeed, it is unlikely that *apitxat* Catalan exhibits a truly categorical absence of /z/ (despite dialectological entries that insist on the absence of /z/ in this variety), and instead is more likely, as attested by Ballart (2013) for *xava* Catalan, to exhibit variability that is socially and linguistically conditioned.

The selection of intervocalic fricatives in Catalan and Spanish for the present study is motivated by the variable voicing and devoicing of Romance fricatives as "natural" and "unremarkable" processes both historically and synchronically (Hualde and Prieto 2014, p. 111). The voicing of intervocalic /s/ to [z] can be characterized as a product of lenition, modeled within a framework of gestural phonology (cf. Browman and Goldstein 1991) as a reorganizing or even undershooting of glottal gestures (e.g., vocal fold abduction) necessary to restrict voicing for [s] while permitting it for the adjacent vowels. As for the devoicing of /z/, the demands for maintenance of a turbulent airstream for sufficient strident frication and the maintenance of voicing are in aerodynamic opposition, which can be resolved with the loss of voicing (Hualde and Prieto 2014, p. 111; see also Ohala 1983, pp. 201–2). The voicing and devoicing of intervocalic sibilants in Romance (e.g., Latin /kása/ > Old Spanish /káza/ > Modern Spanish /kása/ (Penny 2002, pp. 98–103)) accordingly constitute variable processes that are each equally endogenously motivated in Catalan and Spanish, which facilitates the assessment of potential differences in the directionality and asymmetry of contact influence in the present case of Catalan–Spanish contact as all the more reflective of non-linguistic (i.e., social) factors.

4. Research Methodology

The subject population for this study consists of 96 Catalan–Spanish bilinguals, stratified equally by each of gender (male vs. female), age (18–30 vs. 45–60), language profile (L1-Catalan vs. L1-Spanish), and community (Barcelona vs. Valencia). This research was approved by the UC-Berkeley IRB, under protocol # 2016-06-8891. Following the Variationist Sociolinguistic framework (Labov 2001; Tagliamonte 2012), gender stratification, wherein female speakers are likely to use variants with overt negative social stigma less than their male counterparts in cases of stable variation or ongoing change from above[4], is a social constraint that is highly relevant for investigating L1 and L2 differences in the use of an overtly proscribed variant (as is the case for each of Catalan /z/ and Spanish /s/). Along the same vein, age is included in order to assess potential change in progress via generational differences via the apparent time construct (Bailey 2004; Ballart et al. 1991; Chambers 2004). Notably, when applying this methodological construct, patterns of social stratification (especially age and gender) observed in synchronic data are interpreted to evidence possible diachronic trends (i.e., eventual language change which, in the present, is characterized as a potential change in progress), with the understanding that

[4] Changes from above and changes from below, following Labov (2001, pp. 272–74, 279), respectively refer to the community-wide, gradual adoption of a linguistic variant that either is or is not overtly proscribed. Accordingly, the adoption of Spanish [s] and/or Catalan [z] would constitute a change from above, whereas the adoption of Spanish [z] and/or Catalan [s] would constitute a change from below.

"not all variability and heterogeneity in language structure involves change; but all change involves variability and heterogeneity" (Weinreich et al. 1968, p. 188).

With regard to language profile, participants in the present study are grouped according to first language (matched with parents' L1 and the language in the home so as to avoid complications with using the labels "L1" and "L2" with early simultaneous bilinguals (e.g., L1A–L1B)) and self-reported current estimates of typical language use, since, as was previously discussed, functional or practical bilingualism in both languages is widespread in both communities. Table 1 displays the general distribution of the 96 speakers recruited for this study.

Table 1. Subject population.

Language Profile	Speaker Counts (6 Older Male; 6 Older Female; 6 Younger Male; 6 Younger Female)	Home/Native/Parent Native Language	Weekly Use of Catalan (with Friends, Family, School/Work, Shopping)
L1-Catalan/L2-Spanish (Barcelona)	24	Catalan	85% (SD = 7.8)
L1-Spanish/L2-Catalan (Barcelona)	24	Spanish	25% (SD = 9.9)
L1-Catalan/L2-Spanish (Valencia)	24	Catalan	44% (SD = 9.5)
L1-Spanish/L2-Catalan (Valencia)	24	Spanish	8% (SD = 5.2)

Five test instruments were administered to each of the 96 participants. The first test instrument is a sociodemographic questionnaire containing 22 questions used to screen participants according to the social criteria outlined in Table 1.

The second and third test instruments are a pair of recorded phrase-list readings in Catalan and Spanish that elicit self-monitored speech. In each language, subjects were asked to read aloud, using their best Catalan or Spanish pronunciation, a series of 60 target words (all cognates across the languages) with intervocalic Spanish /s/, intervocalic Catalan /z/, and prevocalic word-final Catalan /S/. Target items were stratified according to two linguistic factors across the languages, namely word position (word-medial vs. prevocalic word-final) and syllable stress (unstressed vs. stressed). Word position was included to assess phonotactic variability produced in each language, since the word-medial context in Catalan is the site of phonemic voicing contrast, as opposed to the prevocalic word-final context in which voicing is the result of phonemic neutralization and anticipatory assimilation. The motivation for the inclusion of syllable stress is grounded in the concept of local hyper-articulation for stressed syllables, or the notion that the speaker may reduce otherwise expected effects of gestural overlap with a neighboring segment across stressed syllables, since these kinds of syllables have longer durations and allow the speaker to better time-articulatory gestures independently of one another (Browman and Goldstein 1991; Hualde 2014, p. 251). More concretely, this would suggest that fricative tokens in a syllable with nuclear stress would be the most resistant to voicing as an effect of the greater opportunity (across stressed syllables) for the successful coordination of vocal fold abduction for voiceless [s] relative to the vocal fold adduction gesture of the adjacent nuclear vowel. Token stratification according to word position and stress yielded four cells (word-medial, stressed: *hombre casado/home casat* 'married man'; word-medial, unstressed: *cosa gigante/cosa gigant* 'huge thing'; prevocalic word-final, stressed: *compras agua / compres aigua* 'you buy water'; prevocalic word-final, unstressed: *las amigas/les amigues* 'the friends') of 15 tokens each (per language), which were mixed amongst a set of 60 distractor tokens in each language that did not contain intervocalic fricatives.

The fourth and fifth test instruments consist of a pair of 20-min sociolinguistic interviews in each of Catalan and Spanish, in which participants were asked to discuss their opinions on questions of language identity, the status of Spanish and Catalan in their communities, and issues of linguistic vitality for each language. The interviews accordingly elicited attitudinal data to corroborate sociolinguistic and sociopolitical differences between Catalan varieties in the two communities of study.

Each participant was recorded individually during one experimental session lasting approximately one hour. In order to limit the effects of language mode (Grosjean 2001), given that bilinguals produced Spanish and Catalan speech during a single interview session, the interview session was strictly divided in two parts, namely an L1 portion followed by an L2 portion. The sociodemographic questionnaire was given in each participants' L2, after the L1 tasks (interview and subsequent word reading) and before the L2 tasks (interview and subsequent word reading), providing a buffer of approximately 15 min between language tasks to allow participants to switch from their L1 to their L2. Participants were recorded using an SE50 Samson head-mounted condenser microphone and an H4n Zoom digital recorder (sampling at 44,100 Hz) in an empty classroom at the *Universitat de Barcelona* or *Universitat Pompeu Fabra*, or in a private office at the *Universitat de València*.

Regarding the acoustic analysis of intervocalic fricative tokens, in order to calculate voicing durations for each fricative segment, fricative boundary segmentation was performed manually in Praat by marking left and right boundaries for each segment by using both the waveform and spectrogram to find the zero-intercept in the waveform closest to the first and last signs of aperiodic noise (File-Muriel and Brown 2011, pp. 227–28; Rohena-Madrazo 2015, pp. 298–99). Once intervocalic fricative segments were segmented, exact voicing durations were measured as proportions of each fricative segment that exhibited each of a fundamental frequency (that is, a pitch track), a voice bar at the bottom of the spectrogram, and glottal pulses, with the viewing window exactly twice the size of and centered on the fricative segment (Campos-Astorkiza 2014, p. 21; Gradoville 2011; Hualde 2014, pp. 48–53; Rohena-Madrazo 2015, pp. 298–99; Schmidt and Willis 2011, p. 6; Torreira and Ernestus 2012).[5] Example spectrograms illustrating less voiced and more voiced realizations of intervocalic fricative tokens in Catalan and Spanish appear as Figures 1–4.

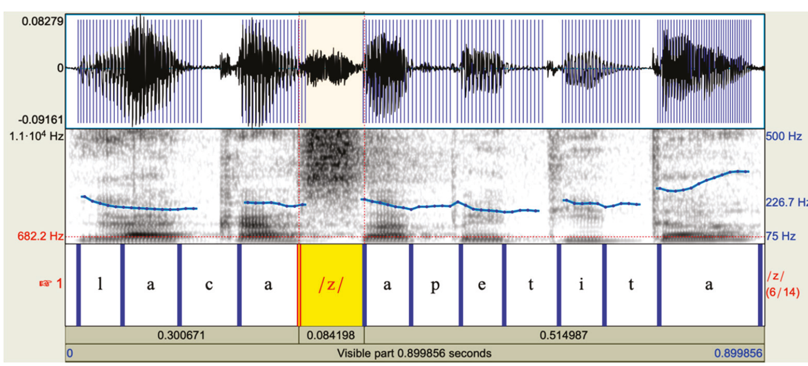


Figure 1. Younger L1-Catalan female rendition of *la ca/z/a petita* ('the little house') in Valencian Catalan (~9% voiced).

---

[5] The manual calculations of segments' proportions of voicing were verified with Praat's voice report automated algorithm, though gross discrepancies between the manual calculation and voicing report were resolved in favor of manual calculation, following Gradoville (2011, pp. 69–71).

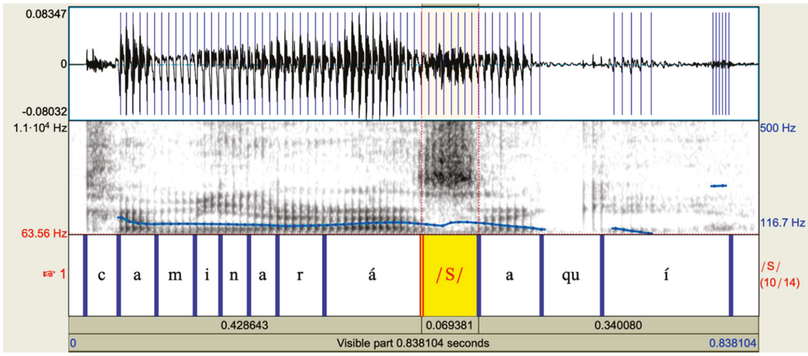


Figure 2. Younger L1-Spanish male rendition of *caminarà/S/ aquí* ('you will walk here') in Barcelonan Catalan (~100% voiced).

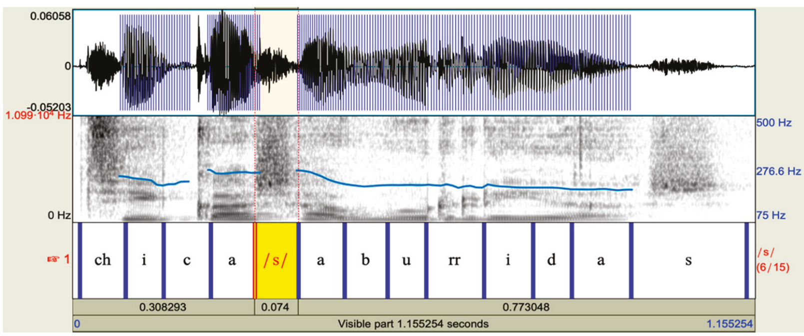


Figure 3. Younger L1-Spanish female rendition of *chicas aburridas* ('bored girls') in Valencian Spanish (~6% voiced).

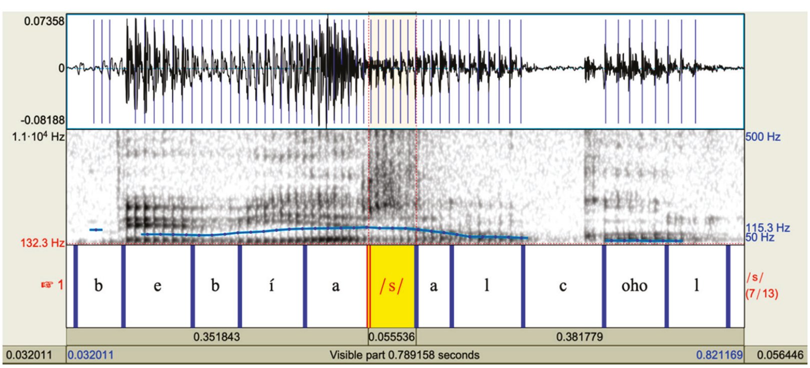


Figure 4. Older L1-Catalan male rendition of *bebía/s/ alcohol* ('you drank alcohol') in Barcelonan Spanish (~100% voiced).

The phrase-list reading tasks in Catalan and Spanish each elicited 5760 intervocalic fricative tokens, yielded 11,520 tokens in total. The relatively few tokens with notable speaker disfluencies (principally pauses between words for prevocalic word-final fricatives) were discarded from analysis, leaving 5654 Catalan tokens and 5635 Spanish tokens. A kernel density plot of all fricatives' voicing proportions per language appears in Figure 5, which evidences a bimodal distribution of voicing proportions.

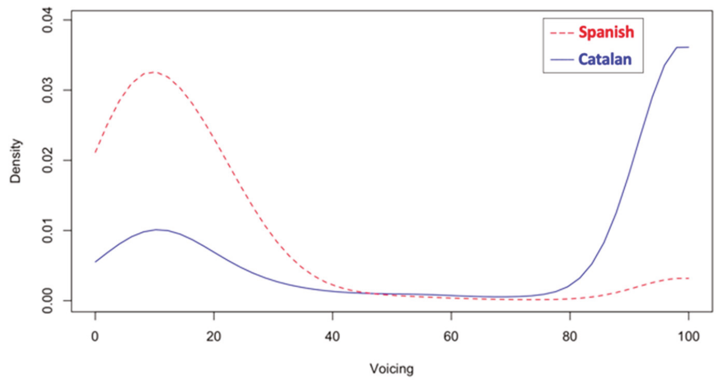


Figure 5. Kernel density plot of Catalan and Spanish intervocalic fricative voicing proportions.

A test of bimodality was conducted in R (R Core Team 2020) using the 'modes' package, which calculates a bimodality coefficient for each language's distribution of voicing proportions ranging from 0 (completely unimodal) to 1 (completely bimodal), for which coefficients greater than 0.555 indicate a bimodal distribution, and coefficients less than or equal to 0.555 indicate a unimodal distribution. The coefficients for Catalan and Spanish fricatives were, respectively, 0.736 and 0.638, indicating bimodal distributions in both languages favoring productions with voicing proportions near either 0% or 100%, with significantly fewer in the middle range of the proportional continuum. Interpreting these modes as articulatory targets for either voiceless [s] or voiced [z], the data were subsequently coded categorically as either [s] for voicing proportions within the range of 0% through 20%, or [z] for voicing proportions within the range of 80% through 100%. This categorical treatment of the bimodal voicing distributions, in line with Campos-Astorkiza (2014), yielded a grand total of 4732 Catalan fricatives and 4578 Spanish fricatives for subsequent statistical analysis (or ~49 Catalan tokens and ~48 Spanish tokens per speaker).

5. Results

5.1. Intervocalic Alveolar Fricative Production

Two mixed-effects logistic regression models (one for Barcelonan data and one for Valencian data) were performed in R (R Core Team 2020) using voicing ([s] vs. [z]) as the dependent variable with treatment contrasts, testing for fixed effects of three linguistic factors (language (Spanish vs. Catalan), word position (medial vs. pre-vocalic word-final), and stress (stressed vs. unstressed)) and three social factors (language profile (L1-Spanish vs. L1-Catalan), gender (male vs. female), and age (older vs. younger)). Interaction terms between language profile, language, and each of all the other independent variables were included in order to assess if any of the remaining effects varied significantly according to language and/or whether the language was the L1 or L2 of each speaker. Individual speaker and token (or word) were included as random effects in both models, for which the alpha level was manually adjusted to 0.025 in order to compensate against Type I errors.

The results of each logistic mixed-effects regression appear in Tables 2 and 3 (note that positive and negative β coefficients respectively indicate greater or lesser log-odds of [z] production relative to the intercept). Given the complex nature of these models, I shall elaborate on them separately, offering additional information and post-hoc analyses as necessary for each finding.

Table 2. Summary of the mixed-effects logistic regression model fitted to Barcelonan fricatives.

	β (in Logits)	*z*	*p*
(Intercept) *	−4.913	−7.967	<0.0001
Catalan	4.882	7.493	<0.0001
Prevocalic word-final	3.751	5.628	<0.0001
Unstressed	2.959	4.791	<0.001
L1-Catalan	−0.145	−0.175	0.893
Younger	3.284	6.174	<0.0001
Female	3.349	6.234	<0.0001
Catalan: Prevocalic word-final	−3.697	−4.913	<0.001
Catalan: Unstressed	0.271	0.285	0.794
Catalan: L1-Catalan	2.172	4.568	<0.001
Catalan: Younger	0.125	0.131	0.859
Catalan: Female	0.163	0.155	0.831
L1-Catalan: Prevocalic word-final	0.752	2.819	0.004
L1-Catalan: Unstressed	0.136	0.189	0.815
L1-Catalan: Younger	0.123	0.174	0.861
L1-Catalan: Female	0.137	0.163	0.769
Catalan: L1-Catalan: Younger	−3.205	−4.638	<0.0001
Catalan: L1-Catalan: Female	−3.403	−4.956	<0.0001
Catalan: L1-Catalan: Prevocalic word-final	−0.028	-0.104	0.927
Catalan: L1-Catalan: Unstressed	−3.003	−4.472	<0.0001

* The intercept is older, L1-Spanish males producing stressed, word-medial fricatives in Spanish.

Table 3. Summary of the mixed-effects logistic regression model fitted to Valencian fricatives.

	β (in Logits)	*z*	*p*
(Intercept) *	−4.924	−7.986	<0.0001
Catalan	3.425	5.736	<0.0001
Prevocalic word-final	2.936	5.148	<0.0001
Unstressed	2.896	4.887	<0.001
L1-Catalan	0.172	0.168	0.884
Younger	−0.184	−0.176	0.831
Female	−0.162	−0.158	0.893
Catalan:Prevocalic word-final	−3.113	−5.472	<0.0001
Catalan: Unstressed	0.165	0.193	0.849
Catalan: L1-Catalan	0.735	3.032	<0.001
Catalan: Younger	−1.049	−4.026	<0.0001
Catalan: Female	−1.472	−4.471	<0.0001
L1-Catalan: Prevocalic word-final	0.685	2.958	<0.001
L1-Catalan: Unstressed	-0.104	−0.162	0.875
L1-Catalan: Younger	−0.116	−0.175	0.858
L1-Catalan: Female	−0.123	−0.182	0.843
Catalan: L1-Catalan: Younger	1.009	4.011	<0.0001
Catalan: L1-Catalan: Female	1.495	4.326	<0.0001
Catalan: L1-Catalan: Prevocalic word-final	−0.152	−0.184	0.874
Catalan: L1-Catalan: Unstressed	0.048	0.115	0.927

* The intercept is older, L1-Spanish males producing stressed, word-medial fricatives in Spanish.

To begin, I focus on attested social constraints on Catalan and Spanish [z] production in Barcelona and Valencia. With respect to language profile, Tukey post-hoc analyses[6] performed on the significant two-way interactions between language profile and language in both communities revealed that while [z] production in Barcelona and Valencia was significantly favored in Catalan over Spanish ($p < 0.0001$

[6] Post-hoc analyses were conducted using the 'emmeans' package (which automatically uses a logit response scale when applied to logistic regression models) with Tukey *p*-value adjustments of multiplicity.

for each community) and by L1-Catalan speakers over L1-Spanish speakers ($p < 0.0001$ for each community), the magnitude of effect for language profile was stronger for Catalan fricatives relative to Spanish fricatives ($p < 0.0001$ for each community). Figures 6 and 7 visualize these differences in Barcelona and Valencia, respectively, and additionally depict the observed categorical favoring of [z] in the Catalan of L1-Catalan speakers in Barcelona. Note that in all subsequent figures, the use of three asterisks denotes comparisons for which $p < 0.0001$, whereas the use of two asterisks denotes comparisons for which $p < 0.001$.

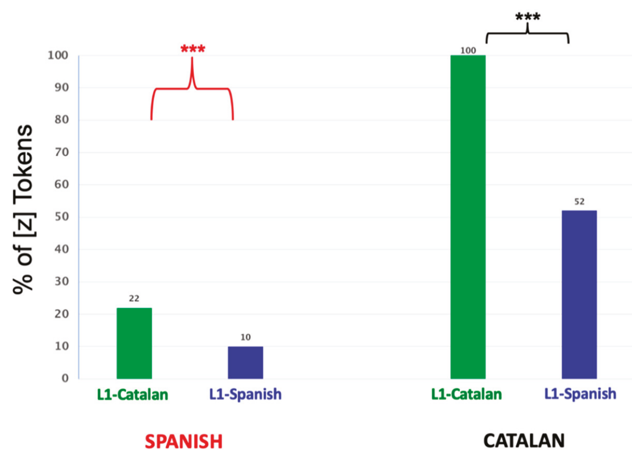


Figure 6. Effects of language profile and language on Barcelonan fricative production.

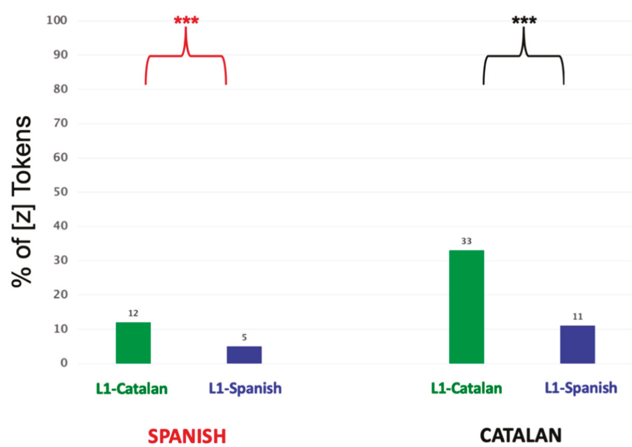


Figure 7. Effects of language profile and language on Valencian fricative production.

With respect to gender, Tukey post-hoc analyses on the significant three-way interaction in Barcelona and Valencia between gender, language profile, and language revealed unique stratifications for each community. For Barcelonan bilinguals, whereas [z] production is favored by females over males in Spanish (for L1-Spanish speakers, $p < 0.0001$; for L1-Catalan speakers, $p < 0.0001$), in Catalan, the parallel gender effect is exclusively present for L1-Spanish speakers ($p < 0.0001$), since L1-Catalan speakers display a categorical favoring of [z] across genders ($p > 0.999$). For Valencian bilinguals, whereas no significant gender stratification is attested in Spanish (for L1-Spanish speakers, $p = 0.749$; for L1-Catalan speakers, $p = 0.768$), in Catalan, [z] is favored by males exclusively for L1-Spanish speakers (for L1-Spanish speakers, $p < 0.0001$; for L1-Catalan speakers, $p = 0.816$). These stratifications are visualized for Barcelona and Valencia in Figures 8 and 9, respectively.

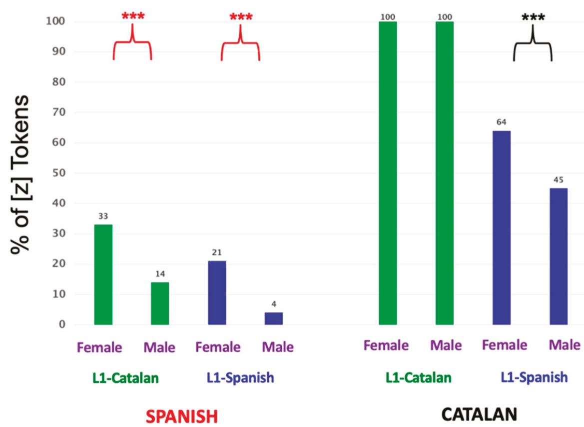


Figure 8. Effect of gender as mediated by language profile and language on Barcelonan fricative production.

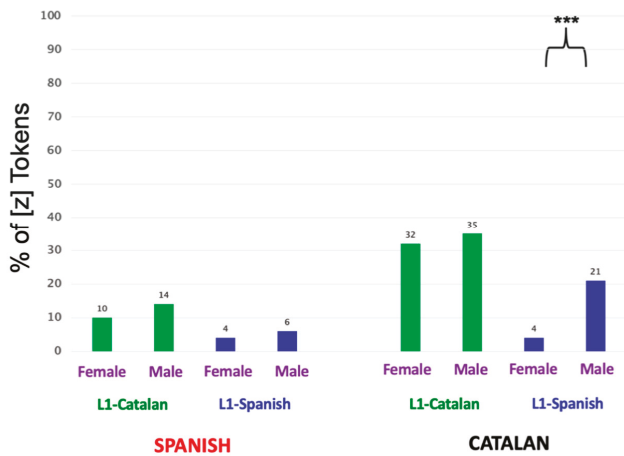


Figure 9. Effect of gender as mediated by language profile and language on Valencian fricative production.

With respect to age, Tukey post-hoc analyses on the significant three-way interaction in Barcelona and Valencia between age, language profile, and language revealed unique stratifications for each community. For Barcelonan bilinguals, whereas [z] production is favored by younger speakers over older speakers in Spanish (for L1-Spanish speakers, $p < 0.0001$; for L1-Catalan speakers, $p < 0.0001$), in Catalan, the parallel age effect is exclusively present for L1-Spanish speakers ($p < 0.0001$), since L1-Catalan speakers display a categorical favoring of [z] across age groups ($p > 0.999$). For Valencian bilinguals, whereas no significant age stratification is attested in Spanish (for L1-Spanish speakers, $p = 0.682$; for L1-Catalan speakers, $p = 0.704$), in Catalan, [z] is exclusively favored by older, L1-Spanish bilinguals (for L1-Spanish speakers, $p < 0.0001$; for L1-Catalan speakers, $p = 0.757$). These stratifications are visualized for Barcelona and Valencia in Figures 10 and 11, respectively.

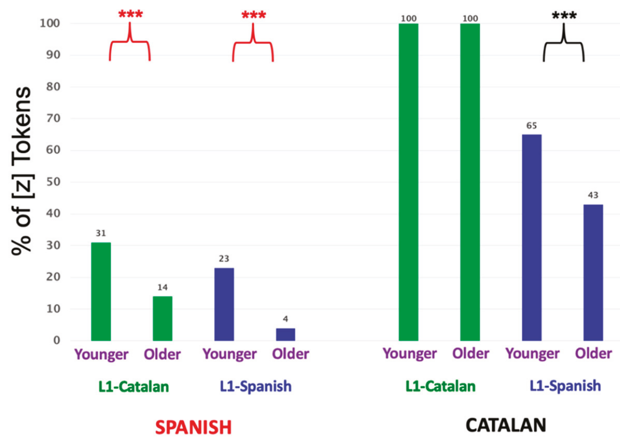


Figure 10. Effect of age as mediated by language profile and language on Barcelonan fricative production.

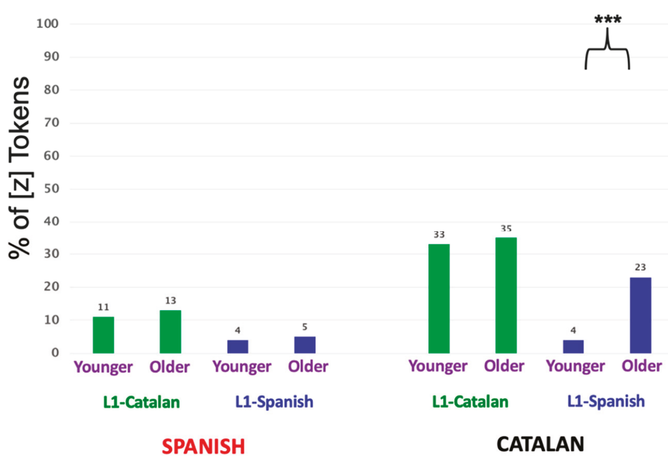


Figure 11. Effect of age as mediated by language profile and language on Valencian fricative production.

As regards linguistic constraints on Barcelonan and Valencian intervocalic fricative production, Tukey post-hoc analyses performed on the pair of significant two-way interactions between word position and each of language profile and language revealed parallel trends in each community. In Barcelona, whereas Spanish [z] production is significantly favored in prevocalic word-final contexts over word-medial contexts (for L1-Spanish speakers, $p < 0.0001$; for L1-Catalan speakers, $p < 0.0001$), Catalan [z] production is not constrained by word position (for L1-Spanish speakers, $p = 0.739$; for L1-Catalan speakers, $p > 0.999$). In Valencia, whereas Spanish [z] production is significantly favored in prevocalic word-final contexts over word-medial contexts (for L1-Spanish speakers, $p < 0.0001$; for L1-Catalan speakers, $p < 0.0001$), Catalan [z] production is not constrained by word position (for L1-Spanish speakers, $p = 0.684$; for L1-Catalan speakers, $p = 0.703$). These stratifications are visualized for Barcelona and Valencia in Figures 12 and 13, respectively, which additionally illustrate the near-categorical absence of [z] tokens in Spanish word-medial contexts across all bilingual participants.

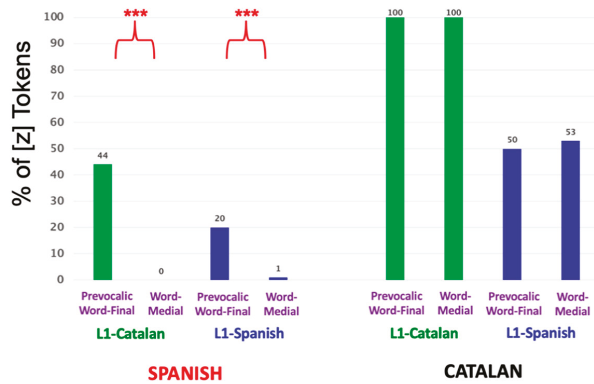


Figure 12. Effect of word position as mediated by language profile and language on Barcelonan fricative production.

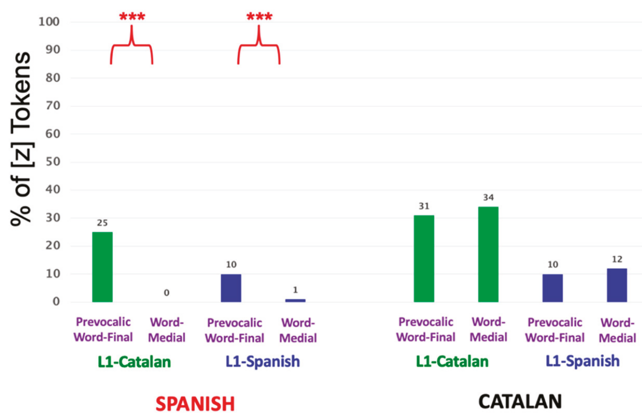


Figure 13. Effect of word position as mediated by language profile and language on Valencian fricative production.

Lastly, with respect to stress, a significant three-way interaction between stress, language profile, and language was attested for Barcelonan bilinguals, whereas a main effect of stress was obtained for Valencian bilinguals. Tukey post-hoc analyses on the significant three-way interaction revealed that whereas [z] production is favored in unstressed contexts over stressed contexts in Barcelonan Spanish (for L1-Spanish speakers, $p < 0.001$; for L1-Catalan speakers, $p < 0.001$), in Barcelonan Catalan, the parallel stress effect is exclusively present for L1-Spanish speakers ($p < 0.0001$), since L1-Catalan speakers display a categorical favoring of [z] across age groups ($p > 0.999$). In Valencia, productions of [z] are similarly favored in unstressed contexts over stressed contexts, independent of language and language profile ($p < 0.001$ for all comparisons). This constraint is visualized for Barcelona and Valencia in Figures 14 and 15, respectively.

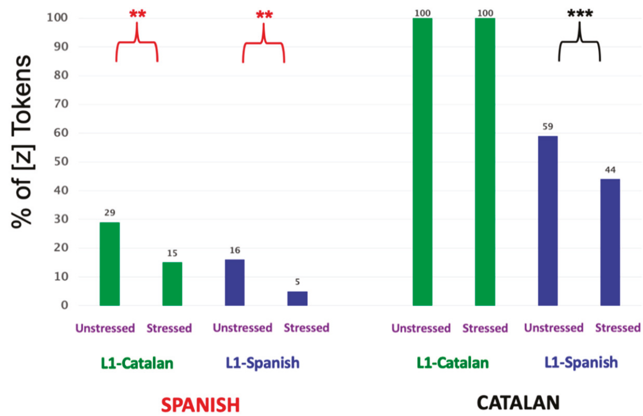


Figure 14. Effect of stress as mediated by language profile and language on Barcelonan fricative production.

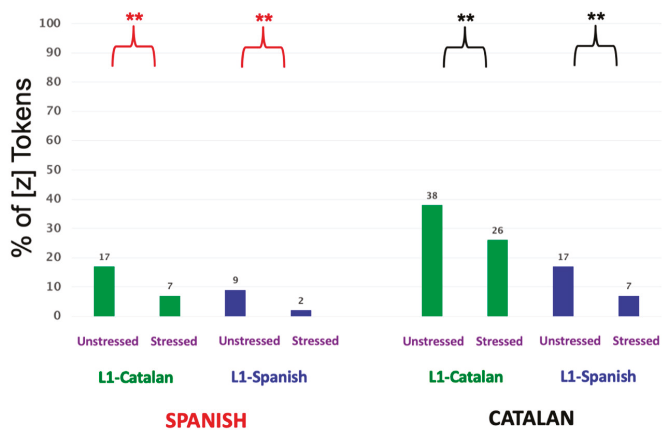


Figure 15. Effect of stress as mediated by language profile and language on Valencian fricative production.

5.2. Language Attitudes

While the majority (95%) of all participants expressed an appreciation of the existence of bilingualism and co-officiality in their respective communities, differences between language attitudes in Barcelona and Valencia largely related to speakers' views toward actively using and promoting Catalan. For example, whereas 81% of Barcelonan participants expressed a desire for their (eventual, hypothetical) children to learn and use Catalan, only 40% of Valencian participants expressed the same desire. As even the Valencian group of L1-Catalan speakers reported a predominance in the use of Spanish over Catalan in their daily lives (refer back to Table 1), perhaps it is unsurprising that the majority (60%) of Valencian participants noted that it was perfectly acceptable to live life in Valencia without even knowing Catalan, and that while it would be nice if their children learned the language, they would not predominantly communicate with them in Catalan.

With regard to language and identity, 100% of Barcelonan participants expressed an association between being Catalan and either understanding the Catalan language or having an appreciation for it. For example, one of the younger female (L1-Catalan) participants noted that "there are many Catalans that choose not to use Catalan, but at least they can understand it and appreciate its presence."

In contrast, 33% of Valencian participants indicated that Valencian identity was tied more strongly to Spanish than to Valencian Catalan, serving to distinguish Valencia from Catalonia: "We put out [on our balconies] Spanish flags and use Spanish to show that in Valencia, we don't reject Spanish like the Catalans do" (Older L1-Spanish Female). Valencian identity as a question of anti-Catalan-ness (via the support of Spanish), rather than as one of Valencian Catalan, is additionally evidenced in the derogatory labeling of overtly pro-Valencian-language individuals as *catalanistas* 'Catalan nationalists': "I've been to Barcelona before, and if you go into a store and speak in Catalan, they either respond in Catalan or in Spanish, but you don't have to change how you speak. Here in Valencia, if you walk into a store speaking Valencian, they'll usually ask you to switch to Spanish, and if you refuse, you're seen as a *catalanista*" (Younger L1-Catalan Male).

When asked if Catalan and Valencian were two different languages, 98% of Barcelonan participants responded negatively, affirming their relationship as related dialects. In Valencia, however, 27% believed Valencian to be an independent language from Catalan. Barcelonan participants were wholly unaware of any conflict regarding the status of Barcelonan Catalan and Valencian Catalan as unique languages, instead noting that Catalan is sometimes wrongly thought to be a dialect of Spanish by outsider, non-Catalonians. Valencian participants, in contrast, were readily able to contextualize the Catalan–Valencian debate within local Valencian politics, noting that it is a point of contention more so for politicians than for the actual Valencian public.

6. Discussion

The patterns of social and linguistic stratification attested for the voiced or voiceless quality of intervocalic fricatives in Barcelonan and Valencian Catalan and Spanish are consistent with unique directionalities and asymmetries of contact influence across these two communities.[7] First, with respect to Barcelonan Catalan and Spanish, evidence in support of Catalan's phonetic influence on Spanish in the form of (prescriptively) non-standard [z] production consists of the observed stratification by language profile, whereby Spanish [z] was favored by L1-Catalan speakers over L1-Spanish speakers. Notably, across both profiles of speaker, Spanish [z] production was nearly categorically constrained by word position, with Spanish [z] appearing nearly singularly in the prevocalic word-final context as opposed to word-medial contexts, the site of phonemic voicing contrast in Catalan. Though the lenition of intervocalic Spanish /s/ (to [h] or [Ø]) in monolingual varieties has similarly been found to be favored word-finally over onset contexts (cf. Hualde and Prieto 2014; Chappell and García 2017; Torreira and Ernestus 2012), the presently observed magnitude of word position effect, categorical for L1-Catalan speakers and near-categorical for L1-Spanish speakers, has not been attested for monolingual Spanish varieties. Moreover, a matched guise study concerning Barcelonan Spanish [z] by Davidson (2019, p. 67) reveals that this feature is covertly associated with Catalan bilingualism within the local bilingual speech community. Taken together with the linguistic stratification by stress (favoring [z] in unstressed contexts), Barcelonan Spanish [z] illustrates a confluence of both endogenous and contact-induced constraints. The additional social stratifications attested for Barcelonan Spanish [z], namely its favoring by younger female speakers, is consistent with a change in progress from below (cf. Labov 2001). In the prevocalic word-final context, younger L1-Catalan females produced [z] at a rate of 74%, which, given the self-monitored nature of the elicited production task, likely undershoots actual [z] production in more casual and spontaneous (or natural) contexts. Accordingly, younger L1-Catalan females lead in the production of Barcelonan Spanish [z] as a majority variant (prevocalic word-finally).

[7] For the assessment of contact effects, I adopt Thomason (2010, 2008, 2001) more flexible treatment of contact-induced innovation as any case in which a linguistic variant is predicted to be more likely to have arisen in the setting of language contact than in a non-contact setting, which is justified or operationalized with respect to sensitivity to specific linguistic and/or social factor constraints consistent with source language agentivity (e.g., a variant's use being mediated by bilingualism and/or language dominance, cognate status with the source language, or any other non-monolingual-like constraint). Language contact accordingly need not be the only (or even principal) source or impetus behind a feature's use in order for it to be considered contact-induced.

With regard to intervocalic fricatives in Barcelonan Catalan, Spanish contact influence can similarly be ascribed through the stratification by language profile, whereby Catalan [s] (in place of prescriptively expected [z] via /z/ and /S/) was favored by L1-Spanish speakers over L1-Catalan speakers. Indeed, Catalan [z] was categorically favored over [s] by L1-Catalan speakers, suggesting that, at least in contexts of more closely self-monitored (or less spontaneous) speech, the phonemic voicing distinction in Catalan is fully maintained. For L1-Spanish speakers, additional social stratifications of age and gender suggest a possible change in progress from above (cf. Labov 2001), with the gradual adoption of more prescriptively normative [z] being led by younger female speakers, who use [z] as a majority variant at a frequency of 68%. Though unconstrained by word-position, voicing rates in Catalan (by L1-Spanish speakers) are greater in unstressed contexts, indicative of the contributions of phonetic-level lenition on the phonological variability of this voicing contrast.

In comparing intervocalic fricative production across Barcelonan Catalan and Spanish, the aforementioned findings illustrate an intriguing asymmetry. Whereas the sociolinguistic stratification of (prevocalic word-final) Spanish [z] indicates an advancing contact variant whose adoption is led by L1-Catalan speakers, the analogous Catalan [s] instead shows signs of gradual abandonment in favor of [z] on behalf of L1-Spanish speakers. Looking at the production frequencies of Spanish (prevocalic word-final) [z] and Catalan [s] by the younger female leaders of each language profile, Spanish [z] is used at over twice the rate of Catalan [s] (respectively, 74% vs. 32%). For younger female L1 speakers of each language, Spanish prevocalic word-final [z] on behalf of L1-Spanish speakers is used at a frequency of 39%, in comparison to Catalan [s] on behalf of L1-Catalan speakers, which is not attested (0%). Accordingly, in Barcelona, the influence of Catalan on Spanish appears considerably stronger than the influence of Spanish on Catalan, though both directions of effect are still present insomuch as both contact variants are favored by L1 speakers of the contact language (i.e., source language agentivity (Van Coetsem 2000)).

With respect to intervocalic fricative production in Valencian Catalan and Spanish, bidirectional contact influence can similarly be observed regarding the usage patterns of Spanish [z] and Catalan [s]. The influence of Catalan on Spanish is attested in the stratification of Spanish [z] by language profile, with L1-Catalan speakers favoring [z] over L1-Spanish speakers. As was the case for Barcelonan Spanish, in Valencian Spanish, across both profiles of speakers, word-position was a near-categorical constraint, effectively barring [z] in the word-medial context, the site of phonemic voicing in Catalan. Unlike in Barcelona, however, no significant social stratifications of age or gender were obtained for Valencian Spanish, which indicates that Catalan [z] (used by L1-Catalan and L1-Spanish speakers, respectively, with frequencies of 25% and 10% in the prevocalic word-final context) is not presently involved in a process of active adoption or change in the community.

As regards the influence of Spanish on Valencian Catalan, Catalan [s] was again favored by L1-Spanish speakers over L1-Catalan speakers, though, in contrast to Barcelonan Catalan, Valencian Catalan [s] is the majority variant even for L1-Catalan speakers, who notably even self-report a greater use of Spanish than Catalan in their daily lives (see Table 1). Additional social stratification in the form of age and gender effects was exclusive to L1-Spanish speakers, favoring Catalan [s] in the speech of younger female speakers, consistent with a change in progress from below (cf. Labov 2001). Younger L1-Spanish females produced [s] at an overall rate of 93%, suggesting that in less self-monitored speech settings, [s] may likely be (near-)categorical, in line with dialectological descriptions of *apitxat* as lacking the voicing contrast (Prieto 2004; Moll 2006). Still, Valencian Catalan [z] was attested roughly one-third of the time by L1-Catalan speakers, underscoring the reality that the *apitxat* variety, like any linguistic variety, is inherently comprised of sociolinguistic variability. Lastly, stress effects favoring [z] production in unstressed contexts (in parallel with Barcelonan Catalan) highlight the role of phonetically lenitive processes in the variability of a phonological voicing contrast.

As was the case for Barcelonan bilinguals, intervocalic fricative production in Valencian Catalan and Spanish evidences crosslinguistic asymmetry. Beyond stratification by language profile, no sociolinguistic correlates were obtained for Spanish [z], whereas for Catalan [s], L1-Spanish younger females lead their older male counterparts in the adoption of this feature. A comparison of usage frequencies between Spanish prevocalic word-final [z] by L1-Catalan speakers (25%) and Catalan [s] by younger L1-Spanish females (93%) illustrates the greater influence (by a magnitude of nearly four) of Spanish on Catalan for this community. For L1-speakers of each language, Spanish prevocalic word-final [z] on behalf of L1-Spanish speakers occurs at a rate of 10%, while Catalan [s] on behalf of L1-Catalan speakers is used at a rate of 67%. Accordingly, in Valencia, the influence of Spanish on Catalan appears considerably stronger than the influence of Catalan on Spanish, though both directions of effect are still present insomuch as both contact variants are favored by L1-speakers of the contact language (i.e., source language agentivity (Van Coetsem 2000)).

Ultimately, the aforementioned findings evidence a case of opposing contact asymmetries across the bilingual communities of Barcelona and Valencia. Operationalized as differential magnitudes between the production of Spanish (prevocalic word-final) [z] and Catalan [s] by L1-Catalan speakers and L1-Spanish speakers, respectively, the influence of Barcelonan Catalan on Barcelonan Spanish is stronger by a factor of approximately two, whereas in Valencia, the influence of Spanish on Catalan is stronger by a factor of approximately four. As the voicing of Spanish /s/ to [z] is just as articulatorily motivated as the devoicing of Catalan /z/ (or /S/) to [s] (Hualde and Prieto 2014, p. 111), differences in the strength of directionality between Catalan as a minority language and Spanish as a majority language can be more transparently linked to the distinct social realities of each language in each community. In Barcelona, the present sociolinguistic interview data corroborate prior claims (cf. Siguan 1988; Sinner 2002) that Catalan is in a position of equal (if not greater) linguistic and social capital than Spanish. Barcelonan speakers in the present investigation readily articulated their esteem of Catalan as part of an expressly bilingual Catalonian identity (corroborating covert attitudes to the same effect in Davidson (2019)), with most advocating for its continued maintenance (if not predominance) amongst subsequent generations of Catalonians. For Catalonians, the active adoption of Spanish [z] is accordingly a " … linguistic resource available to [speakers] in their variety of Spanish as another ethnolinguistic and ideological assertion besides language choice" (Vann 2007, p. 271), the directionality of which (i.e., the greater adoption of Spanish [z] than Catalan [s]) notably mirrors the community's active ideological embrace of Catalan.

In Valencia, in contrast, speakers in the present investigation largely expressed a general apathy toward the use and preservation of Valencian Catalan, tied in part to a social stigma of being too pro-Catalan. The predominant outlook toward Valencian as not a particularly essential language for normal life in Valencia, when coupled with the sizeable minority (33%) of informants that affirmed Spanish as the primary language expressive of Valencian identity, accordingly patterns with the directionality favoring the adoption of Catalan [s] over Spanish [z]. While I do not claim the asymmetry regarding the greater social stratification and use of Valencian Catalan [s] as compared to Valencian Spanish [z] to be a singular, direct consequence of the greater hegemonic distance between Spanish and Catalan in Valencia, the stronger contact influence of Spanish on Valencian Catalan can nonetheless be understood as a probabilistically conditioned outcome of social factors in this community, including population size, sociopolitical status, sociocultural status, and language attitudes (Thomason 2001, 2010; Thomason and Kaufman 1988), all of which uniquely favor Spanish over Catalan in this community. Though both linguistic and social factors are posited to contribute to language variation and change, the present case study, specifically as concerns two equally endogenously motivated changes (e.g., Spanish [z] and Catalan [s]), notably demonstrates how unique social contexts serve to probabilistically favor distinct linguistic outcomes.

7. Conclusions

The present study aimed to explore intervocalic fricative production as a variable feature of Catalan–Spanish contact in two unique communities of Catalan–Spanish bilingualism in order to address questions of directionality and asymmetry of contact influence between them. The unique asymmetries of influence between Catalan and Spanish across Barcelona and Valencia were linked to the asymmetric sociopolitical and sociolinguistic relationships between the languages in each community, which probabilistically condition contact influence at the level of the greater speech community. Accordingly, the social context of language contact plays an essential role in the dynamics of linguistic variation and change in contact settings, in addition to the linguistic and cognitive factors often investigated regarding contact effects at the level of the individual bilingual speaker.

Funding: This research was funded by an Invitational Research Grant (*Atracció de Talent*) sponsored by the *Vicerectorat d'Investigació i Política Científica* (*Servei d'Investigació*) of the *Universitat de València*.

Acknowledgments: This research would not have been possible without the generous hospitality and support provided by Ferran Robles and Maria Labarta Postigo (*Universitat de València*), Antonio Torres Torres, Gemma de Blas (*Universitat de Barcelona*), and *Mireia Trench-Parera* (*Universitat Pompeu Fabra*), as well as Clara Cervera and the Martori family. I am additionally grateful for the helpful comments and insights provided by two anonymous reviewers, in addition to feedback from the members of the UC Davis Symposium on Language Research.

Conflicts of Interest: The author declares no conflict of interest. The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

Acadèmia Valenciana de la Llengua. 2006. *Gramàtica Normative Valenciana*. Valencia: Acadèmia Valenciana de la Llengua.

Bailey, Guy. 2004. Real and apparent time. In *The Handbook of Language Variation and Change*. Edited by Jack K. Chambers, Peter Trudgill and Natalie Schilling-Estes. Malden: Blackwell, pp. 312–32. [CrossRef]

Ballart, Guy, Tom Wikle, Jan Tillery, and Lori Sand. 1991. The apparent time construct. *Language Variation and Change* 3: 241–64. [CrossRef]

Ballart, Jordi. 2013. Variació fònica al català de Barcelona: L'accent xava. *Treballs de Sociolingüística Catalana* 23: 133–51. [CrossRef]

Best, Catherine, and Michael Tyler. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In *Second Language Speech Learning: The Role of Language Experience in Speech Perception and Production*. Edited by Murry Munro and Ocke-Schwen Bohn. Amsterdam: John Benjamins, pp. 13–24. [CrossRef]

Best, Catherine. 1995. A direct realist perspective on cross-language speech perception. In *Speech Perception and Linguistic Experience*. Edited by Winifred Strange. Timonium: York Press, pp. 171–204.

Bosch, Laura, Albert Costa, and Núria Sebastián-Gallés. 2000. First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology* 12: 189–221. [CrossRef]

Browman, Catherine, and Louis Goldstein. 1991. Gestural structures: Distinctiveness, phonological processes, and historical change. In *Modularity and the Motor Theory of Speech Perception: Followings of a Conference to Honor Alvin M. Liberman*. Edited by Ignatius M. Mattingly and Michael Studdert-Kennedy. Hillsdale: Lawrence Erlbaum, pp. 313–38.

Campos-Astorkiza, Rebeka. 2014. Sibilant voicing assimilation in Peninsular Spanish as gestural blending. In *Variation Within and Across Romance Languages: Selected Papers from the 41st Linguistic Symposium on Romance Languages*. Edited by Marie-Hèléne Côté and Eric Mathieu. Philadelphia: John Benjamins, pp. 17–38. [CrossRef]

Carrera-Sabaté, Josefina, Pilar Prieto, and Roser Güell. 2016. Guies de pronunciació del català: Una eina per a l'autoaprentatge de la pronúncia del català. In *L'ensenyament de la pronunciació del català: Eines, recursos, estratègies i experiències*. Edited by Josefina Carrera-Sabaté, Clàudia Pons-Moll and Jesús Bach Marquès. Barcelona: Editorial Graó, pp. 41–56.

Casesnoves Ferrer, Raquel, and David Sankoff. 2004. The Valencian revival: Why usage lags behind competence. *Language in Society* 33: 1–31. [CrossRef]

Casesnoves Ferrer, Raquel. 2010. Changing linguistic attitudes in Valencia: The effects of language planning measures. *Journal of Sociolinguistics* 14: 477–500. [CrossRef]

Chambers, Jack. 2004. Patterns of variation including change. In *The Handbook of Language Variation and Change*. Edited by Jack K. Chambers, Peter Trudgill and Natalie Schilling-Estes. Malden: Blackwell, pp. 349–72. [CrossRef]

Chang, Charles. 2012. Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics* 40: 249–68. [CrossRef]

Chappell, Whitney, and Christina García. 2017. Variable production and indexical social meaning: On the potential physiological origin of intervocalic /s/ voicing in Costa Rican Spanish. *Studies in Hispanic and Lusophone Linguistics* 10: 1–37. [CrossRef]

Davidson, Justin. 2019. Covert and overt attitudes towards Catalonian Spanish laterals and intervocalic fricatives. In *Recent Advances in the Study of Spanish Sociophonetic Perception*. Edited by Whitney Chappell. Philadelphia: John Benjamins, pp. 39–83.

Escudero, Paola. 2005. Linguistic Perception and Second Language Acquisition. Doctoral Thesis, Utrecht University, Utrecht, The Netherlands.

File-Muriel, Richard, and Earl Brown. 2011. The gradient nature of s-lenition in Caleño Spanish. *Language Variation and Change* 23: 223–43. [CrossRef]

Flege, James, and Murray Munro. 1994. The word unit in second language speech production and perception. *Studies in Second Language Acquisition* 16: 381–411. [CrossRef]

Flege, James, David Birdsong, Ellen Bialystok, Molly Mack, Hyekyung Sung, and Kimiko Tsukada. 2006. Degree of foreign accent in English sentences produced by Korean children and adults. *Journal of Phonetics* 34: 153–75. [CrossRef]

Flege, James, Elaina Frieda, and Takeshi Nozawa. 1997. Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics* 25: 169–86. [CrossRef]

Flege, James, Murray Munro, and Ian MacKay. 1995. Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America* 97: 3125–34. [CrossRef] [PubMed]

Flege, James. 1987. The production of 'new' and 'similar' phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15: 47–65. [CrossRef]

Flege, James. 1995. Second-language speech learning: Theory, findings and problems. In *Speech Perception and Linguistic Experience*. Edited by Winifred Strange. Timonium: York Press, pp. 233–77.

Flege, James. 2002. Interactions between the native and second-language phonetic systems. In *An Integrated View of Language Development: Papers in Honor of Henning Wode*. Edited by Petra Burmeister, Thorsten Piske and Andreas Rhode. Trier: Wissenschaftlicher Verlag, pp. 217–44.

Generalitat Valenciana. 2011. *Coneixement i ús del valencià: Dades comparades dels censos de 1986 a 2001*. Valencia: Generalitat Valenciana.

Gradoville, Michael Stephen. 2011. Validity in measurement of fricative voicing: Evidence from Argentine Spanish. In *Selected Proceedings of the 5th Conference on Laboratory Approaches to Romance Phonology*. Edited by Scott Alvord. Somerville: Cascadilla Press, pp. 59–74.

Grosjean, François. 2001. The bilingual's language modes. In *One Mind, Two Languages: Bilingual Language Processing*. Edited by Janet Nicol. Oxford: Blackwell, pp. 1–22.

Guion, Susan, James Flege, and Jonathan Loftin. 2000. The effect of L1 use on pronunciation in Quichua-Spanish bilinguals. *Journal of Phonetics* 28: 27–42. [CrossRef]

Hualde, José Ignacio, and Pilar Prieto. 2014. Lenition of intervocalic alveolar fricatives in Catalan and Spanish. *Phonetica* 71: 109–27. [CrossRef]

Hualde, José Ignacio. 1992. *Catalan*. New York: Routledge.

Hualde, José Ignacio. 2014. *Los sonidos del español*. Cambridge: Cambridge University Press.

Huguet, Ángel. 2006. Small languages and school: The case of Catalonia and the Basque Country. *International Journal of the Sociology of Language* 182: 147–59. [CrossRef]

Institut d'Estadística de Catalunya. 2014. Territori. *Banc d'estadístiques de municipis i comarques*. Available online: http://www.idescat.cat (accessed on 28 March 2019).

Julià i Muné, Joan. 2008. Els sons del català. In *Gramàtica contemporani*. Edited by Joan Solà, Maria-Rosa Lloret and Manuel Pérez Saldanya. Barcelona: Editorial Empúries, vol. 1, pp. 37–87.

Labov, William. 2001. *Principles of Linguistic Change*. Malden: Blackwell.

MacKay, Ian, James Flege, Thorsten Piske, and Carlo Schirru. 2001. Category restructuring during second-language (L2) speech acquisition. *The Journal of the Acoustical Society of America* 110: 516–28. [CrossRef]

Martínez Celdrán, Eugenio, and Ana María Fernández Planas. 2007. *Manual de fonética española*. Barcelona: Ariel.

Moll, Francesc. 2006. *Gramàtica històrica catalana*. Valencia: Universitat de València.

Morgan, Terrell. 2010. *Sonidos en context: Una introducción a la fonética del español con especial referencia a la vida real*. New Haven: Yale University Press.

Navarro Tomás, Tomás. 1918. *Manual de pronunciación española*. Madrid: Imprenta de los Sucesores de Hernando.

Newman, Michael, Mireia Trenchs-Parera, and Shukhan Ng. 2008. Normalizing bilingualism: The effects of the Catalonian linguistic normalization policy one generation after. *Journal of Sociolinguistics* 12: 306–33. [CrossRef]

Ohala, John. 1983. The origin of sound patterns in vocal tract constraints. In *The Production of Speech*. Edited by P. F. MacNeilage. New York: Springer, pp. 189–216. [CrossRef]

Pallier, Christophe, Laura Bosch, and Núria Sebastián-Gallés. 1997. A limit on behavioral plasticity in speech perception. *Cognition* 64: 9–17. [CrossRef]

Penny, Ralph. 2002. *A History of the Spanish Language*. Cambridge: Cambridge University Press. [CrossRef]

Poplack, Shana, and Stephen Levey. 2010. Contact-induced grammatical change: A cautionary tale. In *Language and Space: An International Handbook of Linguistic Variation*. Edited by Peter Auer and Jürgen Erich Schmidt. Berlin: De Gruyter Mouton, vol. 1, pp. 391–419. [CrossRef]

Pradilla, Miquel Ángel. 2001. The Catalan-speaking communities. In *Multilingualism in Spain*. Edited by Maria Teresa Turell Julià. New York: Multilingual Matters Ltd., pp. 58–90. [CrossRef]

Prieto, Pilar. 2004. *Fonètica i fonologia. Els sons del català*. Barcelona: Editorial UOC.

Quilis, Antonio. 1981. *Fonética acústica de la lengua española*. Madrid: Gredos.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, Available online: http://www.R-project.org/ (accessed on 28 March 2019).

Real Acadèmia de Cultura Valenciana. 2000. Estàndart oral valencià. Available online: https://www.llenguavalenciana.com/documents/ortografia/estandart_oral_valencia#sibilants (accessed on 28 March 2019).

Recasens, Daniel. 2014. *Fonètica i fonologia experimentals del català: Vocals i consonants*. Barcelona: Institut d'Estudis Catalans.

Rohena-Madrazo, Marcos. 2015. Diagnosing the completion of a sound change: Phonetic and phonological evidence for /ʃ/ in Buenos Aires Spanish. *Language Variation and Change* 27: 287–317. [CrossRef]

Sanchez, Tara. 2008. Accountability in morphological borrowing: Analyzing a linguistic subsystem as a sociolinguistic variable. *Language Variation and Change* 20: 225–53. [CrossRef]

Schmidt, Lauren, and Erik Willis. 2011. Systematic investigation of voicing assimilation of Spanish /s/ in Mexico City. In *Selected Proceedings of the 5th Conference on Laboratory Approaches to Romance Phonology*. Edited by Scott Alvord. Somerville: Cascadilla Press, pp. 1–20.

Siguan, Miguel. 1988. Bilingual education in Spain. In *International Handbook of Bilingualism and Bilingual Education*. Edited by C. Bratt Paulston. New York: Greenwood, pp. 449–73.

Silva-Corvalán, Carmen. 1986. Bilingualism and language change: The extension of *Estar* in Los Angeles Spanish. *Language* 62: 587–608. [CrossRef]

Sinner, Carsten. 2002. The construction of identity and group boundaries in Catalan Spanish. In *Us and Others: Social Identities across Languages, Discourses, and Cultures*. Edited by Anne Duszak. Philadelphia: John Benjamins, pp. 159–85. [CrossRef]

Tagliamonte, Sali. 2012. *Variationist Sociolinguistics: Change, Observation, Interpretation*. Oxford: Wiley-Blackwell.

Thomason, Sarah, and Terrence Kaufman. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.

Thomason, Sarah. 2001. *Language Contact*. Edinburgh: Edinburgh University Press.

Thomason, Sarah. 2008. Social and linguistic factors as predictors of contact-induced change. *Journal of Language Contact* 2: 42–56. [CrossRef]

Thomason, Sarah. 2010. Contact explanations in linguistics. In *Handbook of Language Contact*. Edited by Raymond Hickey. Somerset: John Wiley & Sons, pp. 31–47. [CrossRef]

Torreira, Francisco, and Mirjam Ernestus. 2012. Weakening of intervocalic /s/ in the Nijmegen corpus of casual Spanish. *Phonetica* 69: 124–48. [CrossRef] [PubMed]

Turell Julià, Maria Teresa. 2000. More on the interplay between internal and external factors: Dialect convergence and divergence of Catalan. In *Dialect and Migration in a Changing Europe*. Edited by Klaus Mattheier. Frankfurt: Peter Lang, pp. 43–56.

Vallverdú, Francesc. 1984. A sociolinguistic history of Catalan. *International Journal of the Sociology of Language* 47: 13–28. [CrossRef]

Van Coetsem, Frans. 2000. *A General and Unified Theory of the Transmission Process in Language Contact*. Heidelberg: Winter Publishing.

Vann, Robert. 1999. Reversal of linguistic fortune: Dimensions of language conflict in autonomous Catalonia. *Language and Communication* 19: 317–27. [CrossRef]

Vann, Robert. 2007. Constructing Catalanism: Motion verbs, demonstratives, and locatives in the Spanish of Barcelona. In *Catalan Review: Sociolinguistics*. Edited by Milton Azevedo, Albert Bastardas, Emili Boix and Paul E. O'Donell. Barcelona: North American Catalan Society, pp. 253–74.

Vila-Pujol, Maria Rosa. 2007. Sociolinguistics of Spanish in Catalonia. *International Journal of the Sociology of Language* 184: 59–77. [CrossRef]

Weinreich, Uriel, William Labov, and Marvin Herzog. 1968. Empirical foundations for a theory of language change. In *Directions for Historical Linguistics*. Edited by Winfred P. Lehmann and Yakov Malkiel. Austin: University of Texas Press, pp. 97–195.

Wheeler, Max. 2005. *The Phonology of Catalan*. Oxford: Oxford University Press.

Winford, Donald. 2005. Contact-induced changes: Classification and processes. *Diachronica* 22: 373–427. [CrossRef]

Winford, Donald. 2010. Contact and borrowing. In *Handbook of Language Contact*. Edited by Raymond Hickey. Somerset: John Wiley & Sons, pp. 170–87. [CrossRef]

Woolard, Kathryn, and Tae-Joong Gahng. 1990. Changing language policies and attitudes in Autonomous Catalonia. *Language in Society* 19: 311–30. [CrossRef]

Woolard, Kathryn. 1984. A formal measure of language attitudes in Barcelona: A note from work in progress. *International Journal of the Sociology of Language* 47: 63–71. [CrossRef]

Woolard, Kathryn. 1989. *Double Talk: Bilingualism and the Politics of Ethnicity in Catalonia*. Palo Alto: Stanford University Press.

Woolard, Kathryn. 2009. Linguistic consciousness among adolescents in Catalonia: A case study from the Barcelonan urban area in longitudinal perspective. *Zeitschrift für Katalanistik* 22: 125–49.

Woolard, Kathryn. 2011. Is there life after high school? Longitudinal changes in the bilingual repertoire in metropolitan Barcelona. *Language in Society* 40: 617–48. [CrossRef]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

languages

Article

Shared or Separate Representations? The Spanish Palatal Nasal in Early Spanish/English Bilinguals

Sara Stefanich [1,*] and Jennifer Cabrelli [2]

1 Department of Spanish and Portuguese, Northwestern University, Evanston, IL 60208, USA
2 Department of Hispanic and Italian Studies, University of Illinois at Chicago, Chicago, IL 60607, USA; cabrelli@uic.edu
* Correspondence: sara.stefanich@northwestern.edu

Received: 3 September 2020; Accepted: 22 October 2020; Published: 2 November 2020

Abstract: The purpose of this study is to examine phonetic interactions in early Spanish/English bilinguals to see if they have established a representation for the Spanish palatal nasal /ɲ/ (e.g., /kaɲon/ *cañón* 'canyon') that is separate from the similar, yet acoustically distinct English /n+j/ sequence (e.g., /kænjn/ 'canyon'). Twenty heritage speakers of Spanish completed a delayed repetition task in each language, in which a set of disyllabic nonce words were produced in a carrier phrase. English critical stimuli contained an intervocalic /n+j/ sequence (e.g., /dɛnjɑ/ 'denya') and Spanish critical stimuli contained intervocalic /ɲ/ (e.g., /deɲja/ 'deña'). We measured the duration and formant contours of the following vocalic portion as acoustic indices of the /ɲ/~/n+j/ distinction. The duration data and formant contour data alike show that early bilinguals distinguish between the Spanish /ɲ/ and English /n+j/ in production, indicative of the maintenance of separate representations for these similar sounds and thus a lack of interaction between systems for bilinguals in this scenario. We discuss these discrete representations in comparison to previous evidence of shared and separate representations in this population, examining a set of variables that are potentially responsible for the attested distinction.

Keywords: heritage bilingualism; early bilingualism; Spanish; English; phonology; phonetics; speech production

---

1. Introduction

An overarching question in the field of bilingual phonology addresses the levels at and degree to which a bilingual's phonetic and phonological systems interact and how these interactions can be modelled within a theory of bilingual grammar. Bilingualism comes in many different forms, one of which is heritage speaker bilingualism. Herein, the term "heritage speaker" (HS) "refer[s] to any bilingual whose [first language] L1 (HL) was learned primarily at home as a minority language and whose [second language] L2 was learned primarily outside the home as the societal (majority) language" (Chang 2020, p. 2). As of result of this acquisition trajectory, heritage speakers' HL and majority language (ML) typically differ with regard to age and context of acquisition, frequency and context of usage, formal education, proficiency, and dominance (which often shifts from the HL to the ML once speakers reach school age), among other factors. These between-language differences yield a unique testing ground for the examination of how these factors modulate the nature of phonetic and phonological interactions in the bilingual mind.

Empirical investigations into the nature and degree of these interactions in heritage speaker phonologies have experienced an uptick over the last decade (see (Chang 2020) for a comprehensive review) and a survey of the growing body of research indicates that production patterns in the heritage language often lie between those attested in late L2 learners of the heritage language that are L1 speakers of the majority language (henceforth, L2ers) and in L1 speakers of the heritage language that have acquired the L2 as adults (henceforth, L1ers). Preliminary evidence suggests that segmental phenomena might be less vulnerable than suprasegmental phenomena to ML influence, albeit with substantial individual variability given the heterogeneity of HSs' language experience. Production data in the ML, on the other hand, although very limited, shows a clearer pattern of production that is typically indistinguishable from that of L1ers, particularly at the segmental level (e.g., Barlow 2014; Mayr and Siddika 2018; McCarthy et al. 2013). In the HL, much of the production research to date has examined segmental phenomena, with attention given to the representation of analogous sounds that are found in monolingual varieties of the ML and HL. That is, researchers have sought to determine whether HSs' production aligns with the baseline production data from L1ers or whether it shows influence from the ML. Existing research varies in outcomes between HL data that skew towards an L1 baseline (e.g., Chang et al. 2009, 2011; Lein et al. 2016) and those that do not (e.g., McCarthy et al. 2013; Ronquest 2012). This variability has been attributed to factors such as speaker generation and sociocultural factors (e.g., Nagy and Kochetov 2013), dominance (e.g., Amengual 2016, 2018; Shea 2019; Simonet 2014), proficiency (e.g., Shea 2019), age of ML acquisition (e.g., Barlow 2014; Cheng 2019), relative similarity between HL and ML sound(s) (e.g., Godson 2003, 2004; Yao and Chang 2016), and whether testing took place in monolingual versus bilingual testing mode (e.g., Amengual 2018; Simonet 2014; Simonet and Amengual 2020), among other things.

Most empirical studies report data solely from heritage speakers' HL, which prevents a direct comparison between HL and ML production that would allow for the verification of distinct HL and ML representations. While comparisons between heritage and baseline data are valuable in their own right, comparisons between the HL and ML are an important contribution to our understanding of heritage speaker phonology in that they allow us to determine the nature of the interaction of the HS's two phonologies. Specifically, we can determine for a crosslinguistic pair of sounds whether a speaker's system includes separate representations utilized in ML versus HL production or a single representation that relies on the production of both the ML and HL. The few studies that report direct comparisons suggest that heritage speakers maintain distinct representations in a shared phonetic space, even when the sound pair under investigation is considered to be similar—but acoustically distinct—in baseline varieties of the HL and ML (e.g., Amengual 2018; Chang et al. 2009, 2011; Knightly et al. 2003).

The studies that have compared HL and ML productions have tested one-to-one analogous sound correspondences between the HL and ML. In the current study, however, we examine a distinct crosslinguistic scenario, specifically the production of nasal sounds in heritage speakers of Spanish in the Midwest US. While the inventory of monolingual Spanish contains the palatal nasal phoneme /ɲ/ (e.g., *cañón* 'canyon' /kaˈɲon/), the inventory of monolingual English does not. However, an approximation exists in the form of the heterosyllabic phoneme sequence /n+j/ (e.g., 'canyon' /ˈkæn.jn/, which can be distinguished acoustically from the complex segment /ɲ/ via the duration and formant trajectories (e.g., Bongiovanni 2019).[1] Herein, we ask whether bilingual speakers of Spanish as the HL and English as the ML rely on distinct representations when producing these sounds in Spanish mode versus English mode.

Data from L1 English/(late) advanced L2 Spanish learners (Stefanich and Cabrelli 2016) have shown that advanced (late) L2 Spanish learners' productions patterned together in English and Spanish modes, and that this apparent shared category did not align with baseline (L1) Spanish data nor with the baseline English data provided by beginner L2 Spanish learners. That is, learners did not appear to create a novel L2 category when producing nonce words that were presented to them auditorily as /ɲ/; Stefanich and Cabrelli (2016) considered this shared intermediate representation to be a potential reflection of L2 influence on an early established L1 representation. This finding aligns with the hypothesis that "similar" sounds in the L2 with an analogue sound in the L1 will be less salient in the

[1] A note on notation: although category representations are often represented in the literature using brackets, we use slashes when referring to phonemic inventories and representations in the speaker's grammar.

input and, in turn, the learner will be less likely to create a novel category for it in the L2 (Flege 1995; Flege and Bohn 2020, but cf., e.g., van Leussen and Escudero 2015, whose (revised) Second Language Linguistic Perception (L2LP) model predicts that similar sounds will be less difficult than different sounds). In light of these L2 data, we examine herein whether early bilinguals' data align with those of their advanced L2 counterparts, or whether these speakers' qualitative and quantitative differences in language experience yield separate representations when in Spanish versus English mode.

After an overview of the relevant nasal consonant inventory in Spanish and English and their acoustic properties, we present the research question and predictions specific to this crosslinguistic scenario. Then, we detail the methods and the results, followed by a discussion. The results from a delayed repetition task administered in separate Spanish and English modes suggest that early bilinguals rely on distinct representations in each mode; the acoustic data indicate that they produce a complex segment in Spanish mode versus a two-segment sequence in English mode. This outcome thus suggests a lack of interaction between systems for these bilinguals in this case, despite the crosslinguistic proximity between English /n+j/ and Spanish /ɲ/. We discuss these discrete representations in comparison to previous evidence of merged versus separate representations in this population and examine the variables that are potentially responsible for the distinction.

1.1. Nasal Consonants in Spanish and English

Spanish has three nasal phonemes that contrast by place of articulation in syllable onset position: bilabial /m/, alveolar /n/, and alveolopalatal /ɲ/ (Díaz-Campos 2004; Recasens 2013) (1).

1. /m/ *cama* /ˈka**m**a/ 'bed';
   /n/ *cana* /ˈka**n**a/ 'gray hair';
   /ɲ/ *caña* /ˈka**ɲ**a/ 'cane'.

The palatal nasal /ɲ/ is the least frequent phoneme in Spanish (Melgar de González 1976) and is a complex segment comprised of an alveolar nasal element followed in succession by a palatal glide element (Martínez Celdrán and Planas 2007; Massone 1988) posited to be phonologically associated with the nasal segment (e.g., Colina 2009).

Although English lacks a phonemic palatal nasal (the inventory is limited to /m n ŋ/), a similar but heterosyllabic /n+j/ sequence is found in words such as *canyon*, *onion*, and *lanyard* (2).

2. *canyon* /ˈkænjn/;
   onion /ˈʌnjn/;
   lanyard /ˈlænjɹ̩d/.

While there are no published data on the acoustic quality of the English /n+j/ sequence to inform the acoustic analysis parameters that distinguish the complex segment /ɲ/ from the discrete segments of /n+j/, a similar (albeit tautosyllabic) sequence is found in Spanish in the surface form of words such as *uranio* /uɾanjo/ 'uranium' and has been investigated acoustically. In Spanish, both /n+j/[2] and /ɲ/ are composed of a combination of a nasal element and a palatal glide element; /ɲ/ is a single complex segment in which the glide element is said to be "partial" (versus a "full" element in /n+j/), Martínez Celdrán and Planas 2007). On the other hand, /n+j/ is a sound sequence in which a "full" glide element is an independent segment. Phonologically, the sound sequence is hypothesized to differ from /ɲ/ in that the glide element in /n+j/ is associated with the following vowel, forming a complex nucleus (e.g., Colina 2009); we assume this to be the case for English /n+j/ as well.

Despite their commonalities, the pair has been found to be distinguished acoustically in word pairs such as *uranio* /uɾanjo/ 'uranium' and *huraño* /uɾaɲo/ 'unsociable' (see Bongiovanni 2019 for a review of acoustic and articulatory evidence). In Bongiovanni's (2019) study of /n+j/ and /ɲ/ production

[2] We employ this phonemic notation following Bongiovanni (2019), recognizing that the glide in this sound sequence in Spanish is not phonemic and that this notation conflates phonetic and phonological representations.

in Buenos Aires Spanish, an analysis of the vocalic portion[3] following the nasal consonant supported the phonological association of the glide to the nasal consonant in /ɲ/ (i.e., ɲV) and the glide to the vowel nucleus in /n+j/ (i.e., njV). Specifically, the gestural difference in the timing and degree of lingual-palatal contact reported in studies such as Recasens (2013) was acoustically evident in formant contour trajectories (i.e., the rise of F2 and the decrease in F1 in /n+j/), the timing at which F1 minimum and F2 maximum were reached (i.e., the timing for /ɲ/ should be earlier), and the duration of the vocalic portion (predicted to be longer in /n+j/ given its status as part of a complex nucleus). Although there is no crosslinguistic research that examines /n+j/ in English versus Spanish, given the heterosyllabic nature of English /n+j/ versus the tautosyllabic /n+j/ in Spanish, it is logical to predict that the glide will be even more clearly associated with the following vowel in English. These predicted differences are visible in the spectrograms and waveforms in Figures 1–3, which are taken from a participant's productions of the nonce item 'denya' in English mode (Figure 1) and 'deña' in Spanish mode (Figure 2), with 'dena' in Spanish mode (Figure 3) as a point of comparison. In the current study, we follow Bongiovanni (2019) and measure duration and formant contours as a correlate of the phonological association of the glide element. As she notes, reporting data from both measures will allow for the confirmation of the reliability of each measure and the avoidance of the overgeneralization of data based on a single measure.

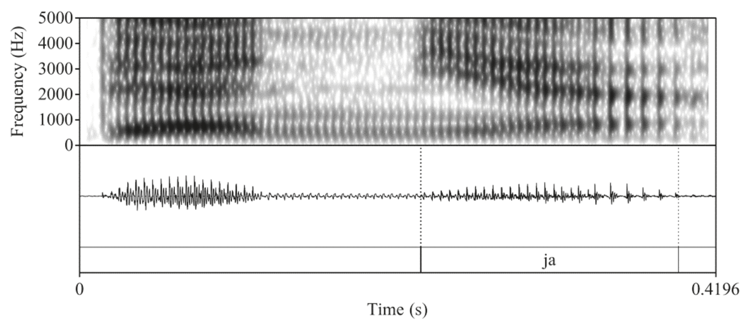


Figure 1. Waveform and spectrogram of a participant's production of /dɛnjɑ/ 'denya' in English mode.

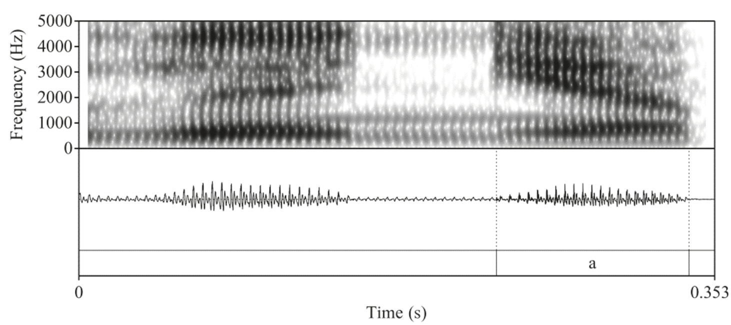


Figure 2. Waveform and spectrogram of a participant's production of /deɲa/ 'deña' in Spanish mode.

[3] In light of the unreliability of acoustic analysis of nasal consonants (see, e.g., Fujimura 1962, cited in Bongiovanni 2019, p. 4), Bongiovanni (2019) limited her analysis to the following vocalic portion.

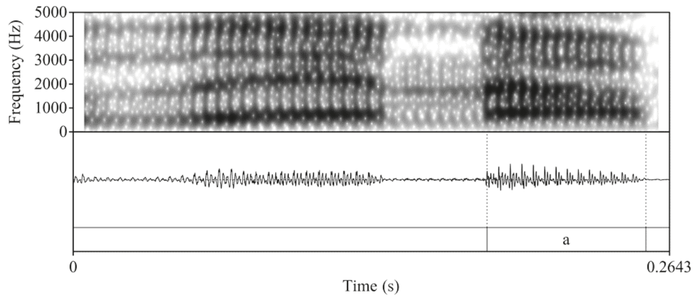


Figure 3. Waveform and spectrogram of a participant's production of /dena/ 'dena' in Spanish mode.

1.2. Research Question and Predictions

The research question that drives this study is the following: do heritage speakers of Spanish evidence distinct representations in their productions of /ɲ/ when in Spanish mode and /n+j/ when in English mode? This is an exploratory question with three possible outcomes: The first is that the differences between language mode in duration and/or formant contours will reveal that these speakers maintain distinct representations. In the case that the quality of these differences patterns with the acoustic parameters associated with /ɲ/ versus /n+j/, such an outcome would be suggestive of implicit knowledge of the distinct single complex segment /ɲ/ in Spanish versus the segment sequence of /n+j/ in English. However, it is wholly possible that speakers will rely on cues other than those reported in the baseline literature, as seen in work on within-language contrasts (e.g., Amengual 2016). The second is that there are no between-language acoustic differences and that the duration and formant contour data skew towards the acoustic description of /n+j/ (i.e., a two-segment sequence rather than a complex segment). The third, like the second, is a lack of between-language differences, but with data that pattern with the acoustic description of /ɲ/. In the latter two cases, it will be necessary to consider what might drive the privileged status of one representation over the other. In terms of predictions, while this is the first study to our knowledge to examine a complex segment compared with a two-segment sequence, we can look to the minimal research that directly compares HL and ML segmental data. As noted in Section 1, when limited to HL data, it is difficult to draw strong conclusions about the interaction of the HL and ML without ML data as a point of comparison. We can of course predict that, if the HL is baseline-like, and knowing that HSs typically are baseline-like in the ML (see Chang 2020, p. 10 for discussion), then they have two representations. Our question, however, is not how close the ML or HL production is to a baseline, but rather whether the speakers' production patterns are different in English mode versus Spanish mode. As we have mentioned, the few studies that directly compare HL and ML data indicate distinct representations in a shared phonetic space, which yields the prediction that the HS in the present study will distinguish acoustically between /ɲ/ and /n+j/ in production. In the case that they do not, we predict that the production will skew towards the reported English pattern, given that (a) the HSs are largely English dominant and (b), overall, segmental differences in the ML when compared to the ML "norm" are small and variable, without clear evidence to date that any measured differences are perceivable (Chang 2020, p. 10).

2. Materials and Methods

2.1. Participants

Twenty Spanish/English bilinguals participated in this study. At the time of the study, all the participants were undergraduate students living in the Chicagoland area. The participants ranged in age from 18–25 (M = 21.05, SD = 1.47). All the participants reported learning Spanish before the age of 3 (M = 0.25, SD = 0.79) and English before the age of 8 (M = 3.30, SD = 2.73). Specifically, six participants reported learning Spanish and English since birth, whereas thirteen reported learning Spanish before

English and one participant learned English before Spanish. We estimate that the majority of these participants are second-generation HS, as approximately 85% of the Spanish HS at the institution where the data were gathered are second-generation speakers (Potowski 2020). The participants reported that, for any given week, they use more English than Spanish with friends and at school/work but more Spanish than English with family (Table 1).

Table 1. Mean percent of language use by domain.

	English		Spanish	
	M	*SD*	*M*	*SD*
Friends	0.82	0.15	0.18	0.15
Family	0.41	0.25	0.59	0.23
School/Work	0.78	0.13	0.20	0.14

As a proxy for language dominance, the participants completed the Bilingual Language Profile (BLP, Birdsong et al. 2012), a bio-linguistic questionnaire which uses the participants' responses to provide a language dominance score on a scale of −218 (Spanish dominant) to 218 (English dominant), with "0" indicating a "balance" between the two languages.[4] The majority of our participants scored on the English side of the scale (n = 17), with a range of scores from −22.7 to 88.6 (M = 43.56, SD = 35.35); the three participants who scored on the Spanish side of the scale fell very close to the balanced zero point. As part of the BLP, the participants rated their Spanish and English proficiency in speaking, understanding, reading, and writing on a scale from 1 (not very well) to 6 (very well) (Table 2). In addition to self-rated proficiency, the participants completed a 50-item written Spanish proficiency assessment composed of portions of the Diploma of Spanish as a Foreign Language (DELE) and Modern Language Association (MLA) assessments commonly administered in heritage research (e.g., Keating et al. 2016; Leal et al. 2015). Our participants averaged a score of 35.50 (SD = 7.80) on the written assessment. Dominance and written proficiency were found to be weakly negatively correlated (r(18) = −0.32, p = 0.175).

Table 2. Self-reported proficiency (scale 1–6).

	English		Spanish	
	M	*SD*	*M*	*SD*
Speaking	5.75	0.55	4.35	1.27
Understanding	5.95	0.22	5.00	0.97
Reading	5.85	0.87	4.05	1.32
Writing	5.75	0.44	3.65	1.27

Heritage speaker populations have been shown to be heterogeneous in terms of language experience and use (e.g., Montrul and Polinsky 2019), and the sample in the current study is no exception. We acknowledge the heterogeneity of these Spanish/English bilinguals in terms of age of acquisition, proficiency, language use, and language dominance and address a number of these factors as they relate to the outcomes in our discussion (Section 4).

2.2. Materials and Procedure

The experiment consisted of Delayed Repetition Tasks (e.g., Trofimovich and Baker 2006) in English mode and Spanish mode. Each task included 40 trials (10 critical, 10 control, 20 distractor). Each trial

[4] Following authors such as Birdsong (2016) and Solis-Barroso and Stefanich (2019), we recognize the gradient nature of the different dimensions of dominance and treat the variable as scalar rather than categorical.

was composed of a target nonce word presented auditorily within the carrier phrase 'I'm saying ___ to you' in English and its equivalent *Digo X para ti* in Spanish. A 1000 ms silent pause was then followed by the spoken prompt "What are you saying to me?" in English or the equivalent *¿Qué me dices?* in Spanish, which prompted the participant to produce the original phrase. Items in both tasks had penultimate stress and were phonotactically licit in the respective language presented. Critical items followed a (C)CV^{1}n.jV^2 (English) or (C)CV^1.ɲV^2 (Spanish) structure and were counterbalanced in each language with 10 control items containing the alveolar nasal /n/ in a (C)CV^1.nV^2 structure.[5] Across critical and control conditions, V^1 was a mid or low vowel (/ɛ/ or /ɑ/ in English and /e/ or /o/ in Spanish; V^2 was /a/ in Spanish and /ɑ/ in English. The 20 distractors followed the same general (C)CV.CV structure as the control and critical items. The item composition in the two tasks is summarized in Table 3; the full set of stimuli is in Appendix A. English stimuli were recorded by a phonetically trained female native speaker of Midwest American English; Spanish stimuli were recorded by a phonetically trained female native speaker of Northern Peninsular Spanish.

Table 3. Composition of nonce stimuli in the English and Spanish delayed repetition tasks.

	n	English	Example		Spanish	Example	
Critical	10	(C)CVn.ja	/dɛnjɑ/ [ˈdɛn.jə]	'denya'	(C)CV.ɲa	/deɲa/ [ˈde.ɲa]	*deña*
Control	10	(C)CV.na	/dɛnɑ/ [ˈdɛ.nə]	'denna'	(C)CV.na	/dena/ [ˈde.na]	*dena*
Distractor	20	(C)CV.CV	/lɛkɑ/ [ˈlɛ.kə]	'lecka'	(C)CV.CV	/meba/ [ˈme.β̞a]	*meba*

Trials were presented using E-prime 2.0 (Psychology Software Tools, Inc., Pittsburgh, PA, USA); audio stimuli were presented over Sennheiser HD-280 PRO (Sennheiser, Wedemark, Germany) headphones through a MOTU Ultralite mk3 interface (MOTU, Cambridge, MA, USA). Recordings took place in a sound-attenuated booth using a head-mounted Shure SM 10A (Shure Inc., Niles, IL, USA) dynamic microphone and a Marantz PMD 661 solid-state recorder (Marantz Corp., Kawasaki, Japan) at a 44.1 kHz sampling rate.

Data were collected in a single session that consisted of separate English and Spanish session modes; the mode order was counterbalanced across participants. All the participants provided informed consent following University of Illinois at Chicago IRB protocol 2015-0040 prior to data collection. The English mode session began with a 10 min interview with the participant to establish the language mode. The participants then completed the English delayed repetition task and the BLP. The Spanish mode session consisted of a 10 min interview, the Spanish delayed repetition task, and the written proficiency assessment.

2.3. Analysis

2.3.1. Acoustic Analysis

Following the literature presented in Section 1.1, this study examines the duration and formant contours of the vocalic portion following the nasal segment. To that end, sound files were segmented and analyzed in Praat [6.1.16] (Boersma and Weenink 2019). The theoretical ceiling of tokens was 600, or 30 per speaker (10 Spanish critical, 10 English critical, 10 Spanish control). One participant's data was excluded from analysis due to a lack of discernible impressionistic difference between /n/ and /ɲ/ in Spanish. Further, an additional 14 tokens were removed due to non-target productions (participants

[5] Spanish alveolar data are reported for contextual comparison; we have excluded the English alveolar data, as they are not relevant to the research question.

skipping, repeating, or producing different segments), creaky voice, or background noise, for a final total of 556 tokens.

During segmentation, the onset of the vocalic portion was determined by the visual presence of an abrupt change in formant structure and frequencies, and the offset was determined by a breaking up of the formant structure and a loss of energy and periodicity in the waveform (Ladefoged 2005). Following Bongiovanni (2019), boundaries between formant transitions or between the glide and the vowel /a/ were not marked.

Once segmentation was completed, measurements were extracted via scripts (Hirst 2012 for automatic duration measurements; McCloy and McGrath 2012 for semi-automatic formant measurements). Formant measurements were taken at 20 points within the vocalic portion (every 5%); 5.2% of the data were manually corrected where it was evident that there were formant tracking errors with the Praat script.

2.3.2. Statistical Analysis

For the duration of the vocalic portion, a linear mixed model (LMM) was fit to the data (measured in ms)[6] using the MIXED procedure in SPSS 26 (IBM Corp. 2019) with a fixed effect of language mode (English, Spanish). The random effects structure (RES) was the maximal structure supported by the data (Barr et al. 2013) and included by-subject and by-item intercepts.

For the formant structure of the vocalic portion, we followed the analysis laid out by Bongiovanni (2019). The formant values were transformed to Bark units, and Smoothing Spline ANOVA (SSANOVA) were fit to the data (time points and corresponding Bark units at each time point) in R, version 4.0.2 (R Core Team 2020), with the gss package. Here, a smoothing spline fits a smooth curve to the observations and the SSANOVA determines whether the curves in question are statistically different from one another (i.e., whether their confidence intervals overlap). As in previous research (e.g., Bongiovanni 2019; Kirkham 2017; Nance 2014; Simonet et al. 2008), we limit our report to the graphical representations of the SSANOVA.

3. Results

3.1. Duration Results

A visual representation of the duration data is presented via the boxplot in Figure 4; as predicted, the vocalic portion of /n+j/ produced in English mode was longer than that of /ɲ/ produced in Spanish mode. The LMM yielded a significant main effect of language ($F(1,41.942) = 70.524$, $p < 0.001$); a Bonferroni post-hoc comparison showed that the vocalic portion for the English /n+j/ ($M = 169.50$ ms, SE = 5.13, CI [159.17,179.84]) was longer than for the Spanish /ɲ/ (107.39 ms, SE = 5.33, CI [96.63,118.16], $p < 0.001$). Hedges' g was calculated as a measure of the effect size on the raw means and standard deviations (English $M = 137.75$ ms, $SD = 42.95$ ms; Spanish $M = 85.78$, $SD = 17.92$) and yielded a large effect size of 1.51 (according to Plonsky and Oswald 2014, for within-subject comparisons). This outcome aligns with the predicted crosslinguistic difference and is indicative of distinct representations in Spanish and English.

[6] To determine the effect of individual differences in speech rate on the outcome, a separate model was fit to z-score-transformed data; the model yielded the same main effect of language ($F(1,41.942) = 70.524$, $p < 0.001$). For ease of interpretation, we report the duration data herein in ms.

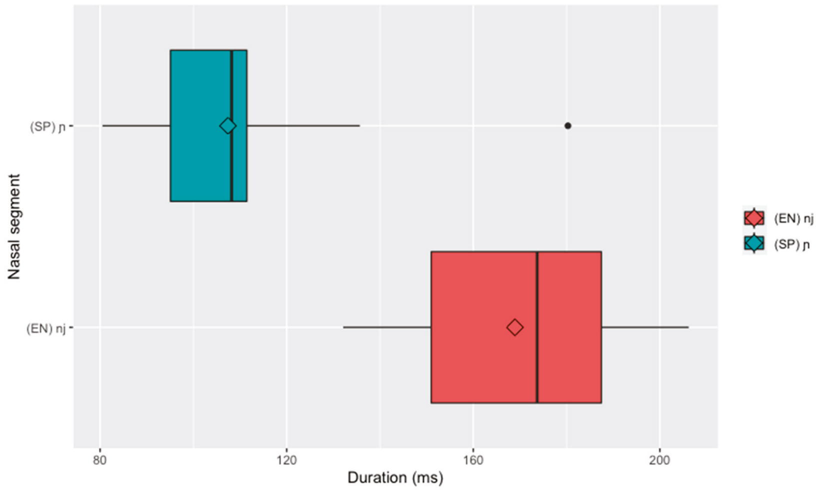


Figure 4. Duration of the following vocalic portion of /n+j/ produced in English mode and /ɲ/ produced in Spanish mode Note: "Nasal segment" refers to /ɲ/ and /n+j/; diamonds represent duration means.

3.2. Formant Structure Results

Recall that, with SSANOVA, statistical significance is indicated by non-overlapping confidence intervals plotted around the data-generated formant curves. The acoustic differences between /ɲ/ and /n+j/ are predicted to take the form of a lower F1 and a higher F2 for /n+j/ than /ɲ/. Keeping these predictions in mind, the results of the SSANOVA are presented in Figure 5.

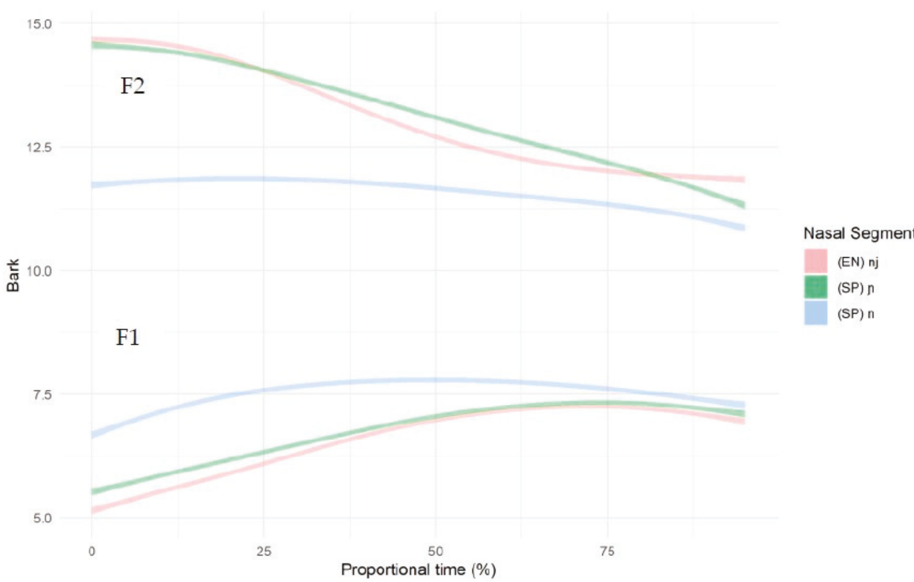


Figure 5. Smoothing Spline ANOVA of formant trajectories by nasal segment.

For F1, the confidence intervals of the /ɲ/ and /n+j/ curves do not overlap between the 0% and 40% points, after which they run adjacent to one another between the 40% and 100% points with a slight overlap between 50% and 75%. For F2, although there is no overlap between 0–20% and 30–80%, the intervals for /ɲ/ and /n+j/ overlap at two points (at roughly 25% and 85%), illustrating a steeper negative slope for /ɲ/ versus /n+j/. These formant readings follow the predicted shapes, with a lower F1 and higher F2 for /n+j/ than for /ɲ/, although these differences are small, with a maximum of between 0.55 and 0.61 Bark at their most different. This difference falls below the assumed just-noticeable difference (JND) threshold of 1 Bark unit, which we address in the discussion in terms of whether this difference is perceivable. In contrast, there is zero overlap in the confidence intervals for Spanish /n/ versus /n+j/ and /ɲ/, with differences that exceed the JND threshold. For Spanish /n/ versus English /n+j/, differences in F1 range from 1.15 to 3.17 Bark and in F2 from 1.61 to 3.68 Bark at their most different. For Spanish /n/ versus /ɲ/, the differences in F1 range from 1.00 to 3.01 Bark and in F2 from 1.48 to 3.16 Bark at their most different.

4. Discussion

4.1. Summary

This study investigated the speech production of a group of Spanish heritage speakers with English as the ML to determine whether their production patterns are acoustically distinct when producing /ɲ/ in Spanish mode versus /n+j/ in English mode. The between-mode differences, which we took to indicate separate representations in a shared phonetic space, were determined via two acoustic indices: (1) the duration of the vocalic portion following the nasal segment (hereafter, FV) and (2) the formant trajectories of the same FV. Acoustic differences were predicted to present in the form of (a) a longer FV duration for /n+j/ than /ɲ/ and (b) formant trajectories in which the /n+j/ evidenced a lower F1 valley and higher F2 peak than /ɲ/, as indicated by non-overlapping formant contours.

The results from the duration analysis confirmed the prediction—the FV for /n+j/ was significantly longer than the FV for /ɲ/. The results from the SSANOVA also fell in line with the expected predictions for the differences between the formant contours; the formant trajectory of the FV for /n+j/ diverged from that of /ɲ/ for portions of the vowel and evidenced a lower F1 and higher F2. Taken together, these results suggest that this group of Spanish HS draws on distinct representations when producing /ɲ/ in Spanish mode versus /n+j/ in English mode. That is, despite the similarities between /n+j/ and /ɲ/, for these participants there is no evidence of interaction between the two phonological systems in this particular case.

4.2. Separate Representations and Age of Acquisition

A lack of interactions between phonological systems suggests that these early bilinguals had sufficient input to develop the representation for the sound they produce when in Spanish mode, despite the fact that /ɲ/ is the least frequent phoneme in the Spanish inventory (Melgar de González 1976). Moreover, they appear to have maintained the representation even after (in most cases) switching dominance to the ML and developing a representation of /n+j/ that is evident in the English mode data. These data align with previous findings that suggest that phonological systems are less susceptible to interaction at the segmental level vs. the suprasegmental level (see Chang 2020 for review). Our findings also add to those that have reported distinct representations of sounds in a shared phonetic space that are similar but acoustically different (e.g., Amengual 2018; Chang et al. 2009, 2011; Knightly et al. 2003, but cf. e.g., Godson 2003; Kang et al. 2016).[7]

[7] One factor that may contribute to why the data do not evidence merged categories, such as those in the voiced stop data in Kang et al. (2016) and the acoustically similar vowel data in Godson (2003), is that some similar crosslinguistic pairs might be "easier" to keep separate. Recall from Section 1.1 that Spanish also has a /n+j/ sequence that contrasts with /ɲ/ in pairs, such as *uranio* /uɾanjo/ 'uranium' and *huraño* /uɾaɲo/ 'unsociable'. Although the only experimental data on this contrast we are aware of is from Buenos Aires Spanish, in which there is a near-merger of /ɲ/ and /nj/, Bongiovanni (2019) found that, even in that case, while the participants did not accurately perceive the difference, their productions were acoustically distinct despite the contrast's low functional load. We posit that one possibility is that the early Spanish bilinguals in this

Interestingly, however, the attested distinction was not found in the production of advanced L1 English/L2 Spanish learners (Stefanich and Cabrelli 2016) in a study that employed the same task. Instead, Stefanich and Cabrelli determined that the advanced L2ers' productions were representative of a merged/hybrid category that was used when producing /ɲ/ in Spanish mode and /n+j/ in English mode. Because these advanced L2ers' data in English mode were different from a group of beginner L2ers' data (used as a proxy for the L2ers' L1 baseline in light of their minimal exposure to L2 input), Stefanich and Cabrelli (2016) posited the development of a hybrid category. While it is important to note that the acoustic index used was the duration of the nasal segment and thus not directly comparable[8], the contrast leads us to the following question: which factors might possibly yield interaction between phonological systems for adult L2ers but not for HSs? A primary difference between these two groups of bilinguals lies in their age of acquisition (AoA).[9] While all the HS participants reported having acquired Spanish before the age of three and English before the age of eight, the mean L2 Spanish AoA was 14.5 (*SD* = 4.21). A substantial body of research suggests a critical period for phonology around 5 years old (e.g., Barlow 2014; Flege et al. 1999; Newport et al. 2001; Scovel 2000, but cf. work that posits a later critical or sensitive period, e.g., DeKeyser 2012, or a lack of one overall, e.g., Abrahamsson and Hyltenstam 2009), and Barlow (2014) indicates that L1 influence on the L2 is more likely after the cutoff age. For instance, in her analysis of Spanish/English laterals, Barlow (2014) found that late bilinguals (AoA > 6) showed evidence of the (English-like) allophonic distribution of [l]~[ɫ] in English and in Spanish compared with early bilinguals (AoA < 5), who only showed it in English. That is, the late bilinguals evidenced interaction between systems (influence from L2 → L1), whereas the early bilinguals did not. Given that this outcome patterns with our HS and L2 results, we suggest that age of acquisition is a good candidate to be a predictor of whether interactions will occur. To confirm this hypothesis, however, we will need to compare groups with early versus late AoA that are matched (as closely as possible) in Spanish proficiency and dominance.

4.3. Individual Variation

While the group results indicate that these heritage speakers have separate representations, there is substantial variability in how this distinction is acoustically realized. While it could be the case that these individual patterns are simply noise in the sample, it is valuable to consider whether certain factors previously reported to condition bilingual speech patterns might explain some of the attested variability. Specifically, we discuss dominance, proficiency, and individual differences related to perception.

Although the purpose of this study was not to *a priori* examine the effects of proficiency and dominance, an analysis of the individual participants' data as they relate to the measures of dominance and proficiency used allows us to examine any trends in the relationship between them and the FV duration and formant contours. Figures 6 and 7 illustrate the difference in duration (in ms) between /n+j/ and /ɲ/ for each participant by proficiency score and dominance score, respectively.

study successfully developed these separate representations early on, and that doing so facilitated the acquisition of the /n+j/ sequence in English. Comparisons of the /n+j/ productions in English versus Spanish mode will inform whether there is a single representation of the /n+j/ sequence or two, thus providing a more complete picture of the crosslinguistic relationship of these similar sounds.

8 Reanalysis of the L2 data from Stefanich and Cabrelli (2016), which will include the measurement of the same acoustic indices (FV duration and formant contours), is in progress.

9 Both groups of bilinguals are overall English dominant, strengthening our conclusion that it is not merely language dominance alone that contributes to the interaction (or lack thereof) between systems. Further, given that language dominance is thought to be fluid and changeable across a bilingual's lifespan (e.g., De Houwer 2011), it makes sense that dominance would not be a determining factor in system interaction.

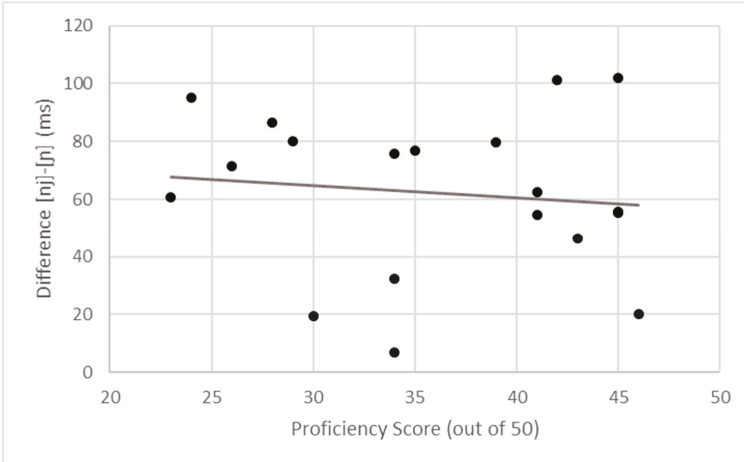


Figure 6. FV duration difference by Spanish proficiency score.

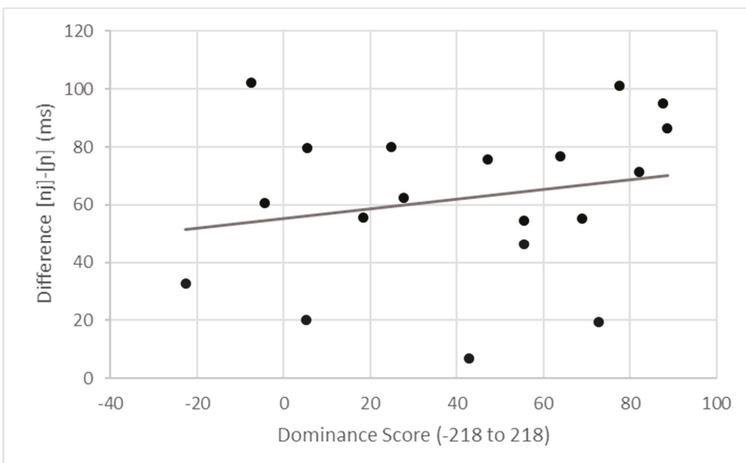


Figure 7. FV duration difference by language dominance score.

Previous research on heritage phonology has found that proficiency and dominance in the HL can at least partially account for individual variation, including the relationship between measures of these constructs and the robustness of a distinction (at least for within-language contrasts—see, e.g., Amengual 2016 for dominance, and Shea 2019, who examined dominance and proficiency and found proficiency to be a stronger predictor). In the current scenario in which we examine between-language distinctions, we might predict that bilinguals with a lower Spanish proficiency score would show smaller differences in FV duration and formant contours than those with higher scores. Regarding dominance, we might expect that bilinguals that fall closer to the balanced zero-point on the dominance scale to evidence greater distinctions than those bilinguals who fall towards the ends of the scale, whose representations would be predicted to skew towards the respective languages.

With respect to the Spanish language proficiency measure reported in Section 2.1, there does not seem to be a discernible pattern; the lack of a relationship visible in Figure 6 is supported by a weak negative correlation (r(18) = −0.11, p = 0.639). If proficiency as measured here played a larger role, we might expect to find a strong positive correlation whereby, as proficiency in the HL increases, so does the difference in duration, maximizing the distance between productions in English versus Spanish modes. This lack of a relationship also seems to be the case for the language dominance measure; the BLP score and duration difference are only weakly correlated ($r(18) = -0.21$, $p = 0.387$). Moreover, if we examine dominance as binary, the three Spanish-dominant participants did not produce durational differences that were substantially larger or smaller than the English-dominant participants, nor did we find any clustering of duration differences around the "balanced" point on the scale.

Turning to the formant contours, the individual SSANOVA (Appendix B), unlike the duration differences, do show differences between participants that can be grouped into four patterns. Of the 19 participants, eight participants' individual splines mimic that of the group-level pattern, i.e., a lower F1 and a higher F2 for /n+j/ than /ɲ/ (Pattern 1). One participant diverges from the group pattern, with /ɲ/ having a lower F1 than /n+j/ (Pattern 2), and seven participants diverge from the group pattern, with /ɲ/ having a higher F2 than /n+j/ (Pattern 3). Lastly, three participants diverge from the group pattern for both F1 and F2. In this latter case, the pattern mirrors that of the group, exhibiting a lower F1 and a higher F2 for /ɲ/ than /n+j/ (Pattern 4). Thus, the majority either align with the group pattern or produce a more fronted[10] FV in /ɲ/ than /n+j/[11]. Figures 8 and 9 present the proficiency and dominance scores by grouped spline pattern to visualize any potential relationships between formant contour patterns and proficiency and dominance scores.

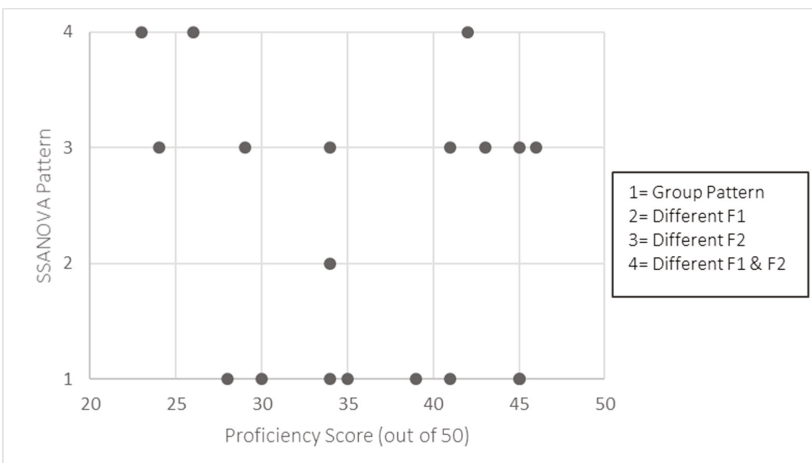


Figure 8. Smoothing Spline pattern type by Spanish proficiency score.

Upon examination of Figures 8 and 9, it does not seem to be the case that either Spanish language proficiency or language dominance as measured in this study can explain the individual patterns. That is, it does not appear that participants with lower Spanish proficiency scores differed from the group pattern more frequently than those with higher scores, or vice versa. Further, the three Spanish-dominant participants did not behave differently from the English-dominant participants. Thus, taking into account the individual analyses for both the FV duration and formant contours,

---

10 As pointed out by an anonymous reviewer, another interpretation of the higher F2 values could be that these speakers are producing a more constricted dorsopalatal realization, given that dorsopalatal constriction narrowing and F2 are positively correlated.

11 As evident in Appendix B, while the individual data fall into the four patterns, there is variation in the degree of spline overlap (i.e., acoustic distance). Without a principled way to quantitatively determine acoustic difference in the formant trajectories, however, we limit our discussion to the categorical patterns and include the individual SSANOVA for readers' reference.

there is no evidence that written Spanish language proficiency score nor language dominance score explain the observed individual variation. Why not? With respect to dominance score, there might not be sufficient variation between participants: recall that the BLP scale is from −218 (Spanish) to 218 (English); our participants' scores are concentrated between −22.7 and 88.6 (M = 43.56, SD = 35.35). Further, we only had three participants with scores on the Spanish side of the scale. Thus, it could be the case, given a wider range of dominance scores and a larger sample, that clearer patterns might emerge with respect to language dominance. Finally, given that dominance can be such an elusive construct, it is possible that a different proxy for dominance could reveal a relationship in the data that the BLP does not; Solis-Barroso and Stefanich (2019) evaluated a set of assessments that were completed by a single group of heritage Spanish bilinguals in Chicago and found different categorization patterns (dominant in Language A, dominant in Language B, "balanced") depending on the assessment, particularly when it came to bilinguals that fell close to the balanced point of a scale.

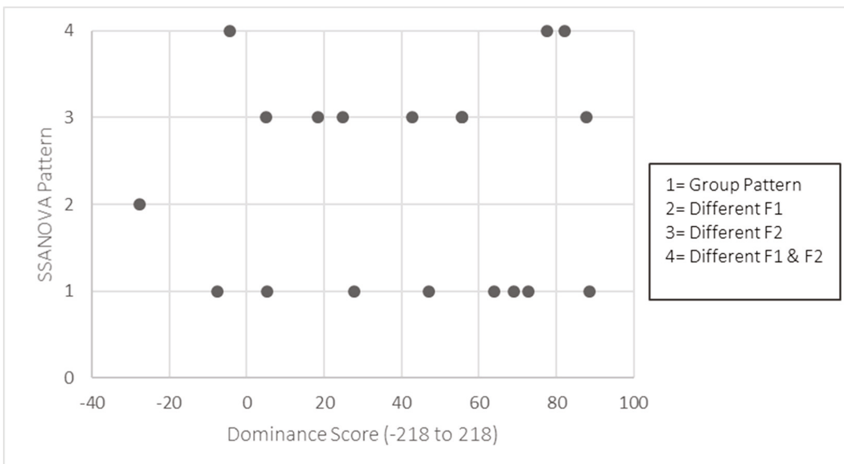


Figure 9. Smoothing Spline pattern type by dominance score.

For the Spanish proficiency score, unlike the dominance score, our participants show a wider range of scores from 23 to 46 (M = 35.50, SD = 7.80), and therefore one might expect to see a difference between those bilinguals with higher scores versus lower scores. However, in our case we see no such difference. Recall that proficiency was measured by a 50-item written measure, which we recognize is not ideal for several reasons, the most relevant of which are that (a) the focus of the study is speech production, and (b) a written assessment disadvantages heritage speakers without substantial formal education in the HL, particularly when the assessment is targeted at speakers of Peninsular Spanish rather than Mexican or US Spanish. Going forward, it would be ideal to employ a measure of oral proficiency (e.g., accent ratings, elicited imitation tests) or a variety of assessments that can be used to formulate a composite score or evaluated independently, such as the set of monologues, picture naming task, and vocabulary assessment used in Shea (2019). It could be the case that, with more direct measures of oral proficiency or a more global evaluation, patterns would more readily emerge in examining interactions between systems.[12]

[12] We also note that we only measured our participants' proficiency in the heritage language (here, Spanish), as our participants attended school and are dominant in the majority language (here English). Future research could also measure proficiency in the majority language in addition to that of the heritage language to see what patterns might surface.

A final consideration when accounting for the attested variation is that of perception. While we have measured variables that have been reported to acoustically distinguish /n+j/ and /ɲ/, it remains to be verified whether these are the acoustic cues that these speakers attend to in the input, even in the cases in which the contour differs beyond the JND threshold. That is, it is wholly possible that these speakers (as a group, or individually) attend to different acoustic cues in the input and that these cues are what they use to distinguish in production, as well. Future research will examine the bilingual perception of the /ɲ/ and /n+j/ in Spanish and English modes, comparing stimuli which vary in the cue(s) (single cues and combination of cues) that are manipulated and held constant. Once we know that these bilinguals perceive the difference between /ɲ/ and /n+j/ and what cues they attend to in the input, an experiment can be designed to isolate production of those cues to confirm separate representations for /n+j/ versus /ɲ/.

4.4. Conclusions

In this paper, we have examined heritage Spanish speakers' crosslinguistic production patterns of /ɲ/ in monolingual Spanish mode and /n+j/ in monolingual English mode to determine whether their phonological systems interact in this scenario. At the group level, we did not find any evidence of interaction and concluded that this group of early Spanish/English bilinguals maintain separate representations for /ɲ/ from /n+j/ based on measures of duration and formant trajectories of the following vocalic portion taken from /ɲ/ data in Spanish mode and /n+j/ in English mode. Comparison with the L2 data suggests that age of acquisition is a likely predictor of interacting systems in this case, at least at the group level. We addressed individual variation in the sample via the relationships between duration and formant contours and dominance and proficiency, and did not find any clear explanatory trends. The next step in this line of investigation will thus be to determine why some bilinguals evidence formant trajectory patterns at the individual level that diverge from the group-level pattern. To that end, we highlighted the need to (a) replicate the study with a larger sample that spans a wider range of proficiency and dominance and (b) test perception to isolate the acoustic cues that are used to distinguish /ɲ/ and /n+j/ in the input. It will also be valuable to directly compare the heritage data with non-heritage native speaker data in the HL to determine how the representations of these populations, who typically differ in dominance and input quantity/quality (among other factors), overlap. Finally, it will be of interest to determine whether language mode plays a role in the interaction of bilinguals' systems when it comes to these sounds. In the current study, we tested in monolingual modes in order to give participants the best chance possible of producing distinct segments. However, research on language mode in HS has shown that a bilingual versus monolingual mode in testing plays a role in both production and perception (e.g., Amengual 2018; Antoniou et al. 2012; Simonet and Amengual 2020). What would happen if were to test these HS in a bilingual mode (which is common for this community, in which participants are able to—and often do—codeswitch between the languages)? Would we see evidence of interaction? If so, what does that tell us about the nature of these representations? Ultimately, the triangulation of data from various bilingual profiles in monolingual and bilingual testing modes will lead us further towards the goal of a holistic understanding of the nature of interacting systems in the bilingual brain.

Author Contributions: Conceptualization, S.S. and J.C.; methodology, S.S. and J.C.; software, S.S. (for data), J.C. (for statistical analysis); validation, S.S. and J.C.; formal analysis, S.S. and J.C.; investigation, S.S. and J.C.; resources, J.C.; data curation, S.S.; writing—original draft preparation, S.S. and J.C.; writing—review and editing, S.S. and J.C.; visualization, S.S. and J.C.; supervision, S.S. and J.C.; project administration, S.S. and J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: We would like to thank David Abugaber for his assistance with the SSANOVA analyses as well as Leire Echevarria and Brian Rocca for their help with data collection. We also wish to thank Mark Amengual as the Guest Editor of this issue and the two anonymous reviewers for their valuable feedback.

Conflicts of Interest: The authors declare no conflict of interest.

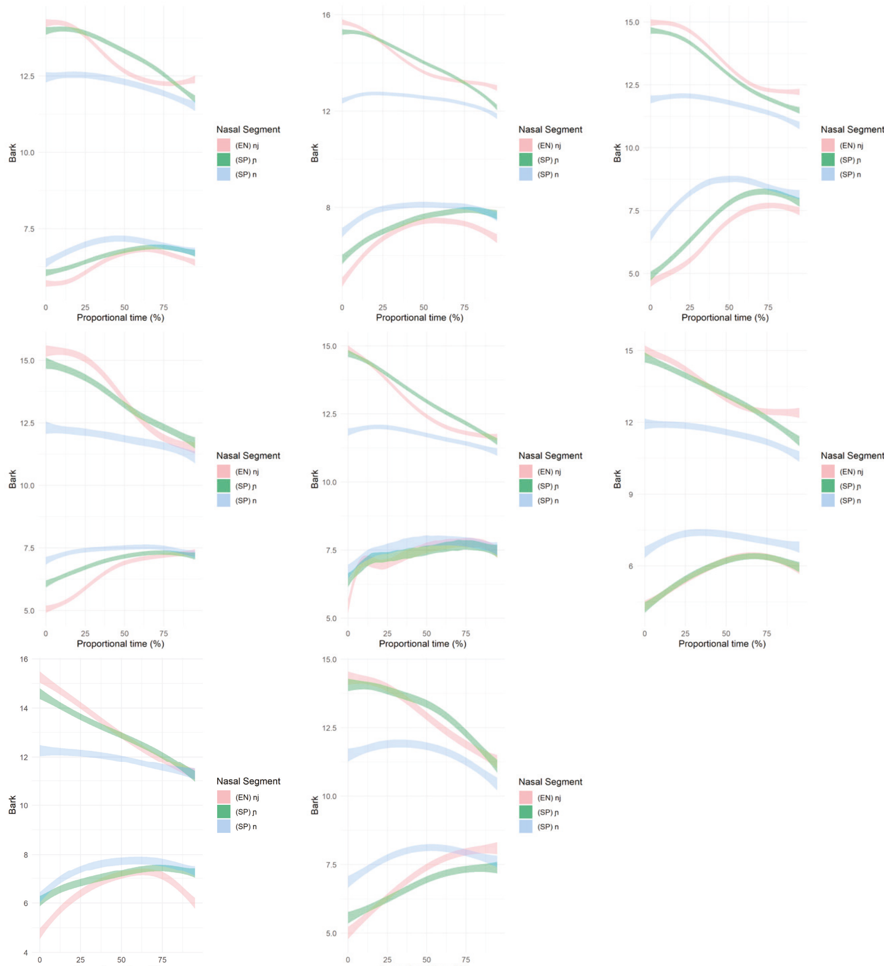
Appendix A

Table A1. Stimuli.

	Spanish Mode		**English Mode**	
Critical (C)CV.ɲa (Spanish) (C)CVn.ja (English)	reña	[reɲa]	renya	[ɹɛnjə]
	boña	[boɲa]	bonya	[bɑnjə]
	broña	[bɾoɲa]	bronya	[bɹɑnjə]
	droña	[dɾoɲa]	dronya	[dɹɑnjə]
	feña	[feɲa]	fenya	[fɛnjə]
	poña	[poɲa]	ponya	[p^{h}ɑnjə]
	foña	[foɲa]	fonya	[fɑnjə]
	loña	[loɲa]	lonya	[lɑnjə]
	deña	[deɲa]	denya	[dɛnjə]
	beña	[beɲa]	benya	[bɛnjə]
Control (C)CV.na	bena	[bena]	benna	[bɛnə]
	dena	[dena]	denna	[dɛnə]
	lona	[lona]	lonna	[lɑnə]
	fona	[fona]	fonna	[fɑnə]
	pona	[pona]	ponna	[p^{h}ɑnə]
	fena	[fena]	fenna	[fɛnə]
	drona	[dɾona]	dronna	[dɾɑnə]
	brona	[bɾona]	bronna	[bɾɑnə]
	quena	[kena]	renna	[ɹɛnə]
	jona	[xona]	bonna	[bɑnə]
Distractor	nela	[nela]	talla	[t^{h}ælə]
	neda	[neð̞a]	tamma	[t^{h}æmə]
	dera	[deɾa]	tulla	[t^{h}ʌlə]
	gada	[gað̞a]	bura	[bɚə]
	meba	[meβ̞a]	lekka	[lɛkə]
	bera	[beɾa]	meppa	[mɛpə]
	doda	[dođ̞a]	maffa	[mæfə]
	bora	[boɾa]	ponka	[p^{h}ɑnkə]
	doba	[doβ̞a]	cromma	[k^{h}ɹɑmə]
	gora	[goɾa]	neppa	[nɛpə]
	gera	[geɾa]	zappa	[zæpə]
	pada	[pað̞a]	ficka	[fɪkə]
	fala	[fala]	vatta	[væɾə]
	deda	[deð̞a]	virta	[vɚɾə]
	seba	[seβ̞a]	zanta	[zæntə]
	poba	[poβ̞a]	thappa	[θæpə]
	dola	[dola]	thurpa	[θɚpə]
	teba	[teβ̞a]	drotta	[dɹɑɾə]
	dela	[dela]	vecka	[vɛkə]
	bada	[bað̞a]	stucka	[stʌkə]

Appendix B

(**a**)



(**b**)

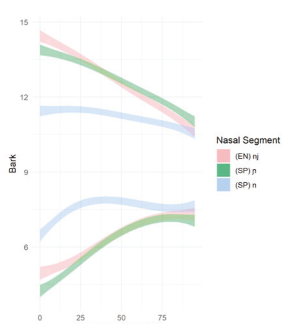
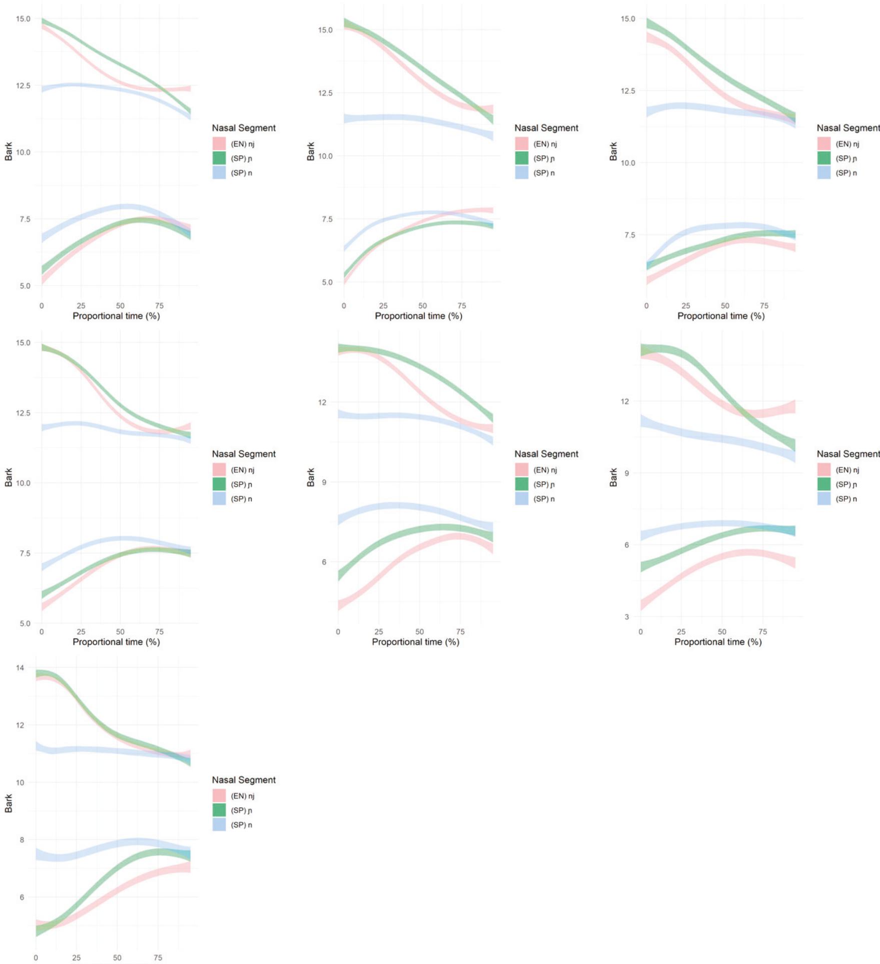


Figure A1. *Cont.*

(**c**)



(**d**)

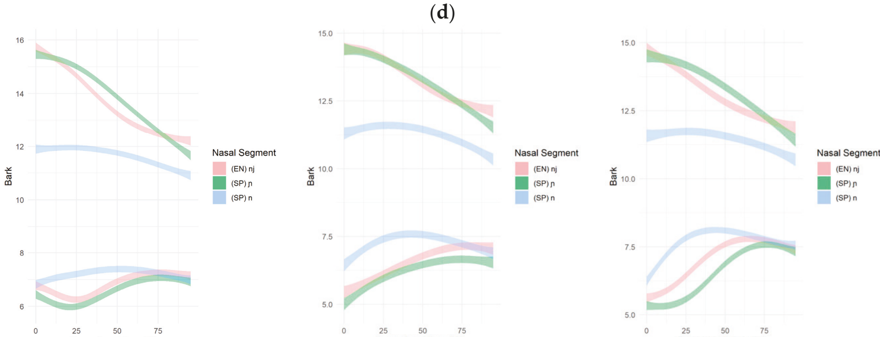


Figure A1. (**a**) Pattern 1: individual splines that follow the group pattern. (**b**) Pattern 2: individual splines that differ in F1 from the group pattern. (**c**) Pattern 3: individual splines that differ in F2 from the group pattern. (**d**) Pattern 4: individual splines that differ in F1 and F2 from the group pattern.

References

Abrahamsson, Niclas, and Kenneth Hyltenstam. 2009. Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning* 59: 249–306. [CrossRef]

Amengual, Mark. 2016. Acoustic correlates of the Spanish tap-trill contrast: Heritage and L2 Spanish speakers. *Heritage Language Journal* 13: 88–112. [CrossRef]

Amengual, Mark. 2018. Asymmetrical interlingual influence in the production of Spanish and English laterals as a result of competing activation in bilingual language processing. *Journal of Phonetics* 69: 12–28. [CrossRef]

Antoniou, Mark, Michael Tyler, and Catherine Best. 2012. Two ways to listen: Do L2 dominant bilinguals perceive stop voicing according to language mode? *Journal of Phonetics* 40: 582–94. [CrossRef] [PubMed]

Barlow, Jessica A. 2014. Age of acquisition and allophony in Spanish-English bilinguals. *Frontiers in Psychology* 5: 288. [CrossRef] [PubMed]

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68: 255–78. [CrossRef] [PubMed]

Birdsong, David. 2016. Dominance in bilingualism: Foundations of measurement, with insights from the study of handedness. In *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*. Edited by Jeanine Treffers-Daller and Carmen Silva-Corvalán. Cambridge: Cambridge University Press, pp. 85–105.

Birdsong, David, Libby M. Gertken, and Mark Amengual. 2012. *Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism*. Austin: COERLL, University of Texas at Austin.

Boersma, Paul, and David Weenink. 2019. Praat: Doing Phonetics by Computer (Version 6.1.22) [Computer Program]. Available online: http://www.praat.org/ (accessed on 31 October 2020).

Bongiovanni, Silvina. 2019. An acoustical analysis of the merger of /ɲ/ and /nj/ in Buenos Aires Spanish. *Journal of the International Phonetic Association*, 1–25. [CrossRef]

Chang, Charles. 2020. Phonetics and Phonology. In *The Cambridge Handbook of Heritage Languages and Linguistics*. Edited by Silvina Montrul and Maria Polinsky. Cambridge: Cambridge University Press. Available online: https://drive.google.com/file/d/1Pbks63JsZwcfyaUgjm-yB1EbwWGcmhil/view (accessed on 31 October 2020).

Chang, Charles, Erin F. Haynes, Yao Yao, and Russell Rhodes. 2009. A tale of five fricatives: Consonantal contrast in heritage speakers of mandarin. *University of Pennsylvania Working Papers in Linguistics* 15: 37–43.

Chang, Charles, Yao Yao, Erin F. Haynes, and Russell Rhodes. 2011. Production of phonetic and phonological contrast by heritage speakers of mandarin. *The Journal of the Acoustical Society of America* 129: 3964–80. [CrossRef] [PubMed]

Cheng, Andrew. 2019. Age of arrival does not affect childhood immigrants' acquisition of ongoing sound change: Evidence from Korean Americans. In *Proceedings of the 19th International Congress of Phonetic Sciences*. Edited by Sasha Calhoun, Paola Escudero, Marija Tabain and Paul Warren. Canberra: Australasian Speech Science and Technology Association Inc., pp. 2213–17.

Colina, Sonia. 2009. *Spanish Phonology: A Syllabic Perspective*. Washington: Georgetown University Press.

De Houwer, Annick. 2011. Language input environments and language development in bilingual acquisition. *Applied Linguistics Review* 2: 221–40. [CrossRef]

DeKeyser, Robert. 2012. Age effects in second language learning. In *The Routledge Handbook of Second Language Acquisition*. Edited by Susan Gass and Alison Mackey. New York: Routledge, pp. 442–60.

Díaz-Campos, Manuel. 2004. Context of learning in the acquisition of Spanish second language phonology. *Studies in Second Language Acquisition* 26: 249–73. [CrossRef]

Flege, James E. 1995. Second language speech learning: Theory, findings, and problems. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Edited by Winifred Strange. Baltimore: York Press, pp. 233–72.

Flege, James E., and Ocke-Schwen Bohn. 2020. The revised speech learning model (slm-r). In *Second Language Speech Learning: Theoretical and Empirical Progress*. Edited by Ratree Wayland. Cambridge: Cambridge University Press, in press.

Flege, James E., Ian R. A. MacKay, and Diane Meador. 1999. Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America* 106: 2973–87. [CrossRef] [PubMed]

Fujimura, Osamu. 1962. Analysis of nasals consonants. *The Journal of the Acoustical Society of America* 34: 1865–75. [CrossRef]

Godson, Linda. 2003. Phonetics of Language Attrition: Vowel Production and Articulatory Setting in the Speech of Western Armenian Heritage Speakers. Ph.D. thesis, University of California, San Diego.

Godson, Linda. 2004. Vowel production in the speech of western Armenian heritage speakers. *Heritage Language Journal* 2: 44–69.

Hirst, Daniel. 2012. Analyse_tier.praat. Available online: http://uk.groups.yahoo.com/group/praat-users/files/Daniel_Hirst/analyse_tier.praat (accessed on 13 October 2015).

IBM Corp. 2019. *IBM SPSS Statistics for Macintosh, Version 26.0*. Armonk: IBM Corp.

Kang, Yoonjung, Sneha George, and Rachel Soo. 2016. Cross-language influence in the stop voicing contrast in heritage Tagalog. *Heritage Language Journal* 13: 184–218. [CrossRef]

Keating, Gregory D., Jill Jegerski, and Bill VanPatten. 2016. Online processing of subject pronouns in monolingual and heritage bilingual speakers of Mexican Spanish. *Bilingualism: Language and Cognition* 19: 36–49. [CrossRef]

Kirkham, Sam. 2017. Ethnicity and phonetic variation in Sheffield English liquids. *Journal of the International Phonetic Association* 47: 17–35. [CrossRef]

Knightly, Leah M., Sun-ah Jun, Janet S. Oh, and Terry kit-fong Au. 2003. Production benefits of childhood overhearing. *The Journal of the Acoustical Society of America* 114: 465–74. [CrossRef]

Ladefoged, Peter. 2005. *A Course in Phonetics*, 5th ed. Boston: Thomson.

Leal, Tania, Jason Rothman, and Roumyana Slabakova. 2015. Discourse-sensitive clitic-doubled dislocations in heritage Spanish. *Lingua* 155: 85–97. [CrossRef]

Lein, Tatjana, Tanja Kupisch, and Joost van de Weijer. 2016. Voice onset time and global foreign accent in german–french simultaneous bilinguals during adulthood. *International Journal of Bilingualism* 20: 732–49. [CrossRef]

Martínez Celdrán, Eugenio, and Ana María Fernández Planas. 2007. *Manual de Fonética Española: Articulaciones y Sonidos de Español*. Madrid: Editorial Ariel.

Massone, Maria Ignacia. 1988. Estudio acústico y perceptivo de las consonantes nasales y líquidas del español [Acoustic and perceptual study of Spanish nasal and liquid consonants]. *Estudios de Fonética Experimental* 3: 13–34.

Mayr, Robert, and Aysha Siddika. 2018. Inter-generational transmission in a minority language setting: Stop consonant production by Bangladeshi heritage children and adults. *International Journal of Bilingualism* 22: 255–84. [CrossRef]

McCarthy, Kathleen M., Bronwen Evans, and Merle Mahon. 2013. Acquiring a second language in an immigrant community: The production of Sylheti and English stops and vowels by London-Bengali speakers. *Journal of Phonetics* 41: 344–58. [CrossRef]

McCloy, Daniel, and August McGrath. 2012. SemiAutoFormantExtractor.praat. Available online: https://github.com/drammock/praat-semiauto/blob/master/SemiAutoFormantExtractor.praat (accessed on 10 February 2018).

Melgar de González, Maria. 1976. *Cómo Detectar al niño con Problemas de Habla [How to Detect a Child with Speech Problems]*. Mexico City: Trillas.

Montrul, Silvina, and Maria Polinsky. 2019. Introduction to heritage language development. In *The Oxford Handbook of Language Attrition*. Edited by Monika Schmid and Barbara Köpke. Oxford: Oxford University Press, pp. 419–33.

Nagy, Naomi, and Alexei Kochetov. 2013. Voice onset time across the generations: A cross-linguistic study of contact-induced change. In *Multilingualism and Language Contact in Urban Areas: Acquisition—Development—Teaching—Communication*. Edited by Peter Siemund, Ingrid Gogolin, Monica Edith Schulz and Julia Davydova. Amsterdam: John Benjamins, pp. 19–38.

Nance, Claire. 2014. Phonetic variation in Scottish Gaelic laterals. *Journal of Phonetics* 47: 1–17. [CrossRef]

Newport, Elissa, Daphne Bavelier, and Helen Neville. 2001. Critcal thinking about critical periods: Perspectives on a critical period for language acquisition. In *Language, Brain and Cognitive Development: Essays in Honor of Jacques Mehler*. Edited by Emmanuel Dupoux. Cambridge: MIT Press, pp. 481–502.

Plonsky, Luke, and Frederick Oswald. 2014. How big is "big"? Interpreting effect sizes in l2 research. *Language Learning* 64: 878–912. [CrossRef]

Potowski, Kim. 2020. University of Illinois at Chicago, Chicago, IL, USA. Personal communication, August 10.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, Available online: https://www.R-project.org/ (accessed on 31 October 2020).

Recasens, Daniel. 2013. On the articulatory classification of (alveolo) palatal consonants. *Journal of the International Phonetic Association* 43: 1–22. [CrossRef]

Ronquest, Rebecca. 2012. An Acoustic Analysis of Heritage Spanish Vowels. Ph.D. thesis, Indiana University, Bloomington.

Scovel, Thomas. 2000. A critical review of the critical period research. *Annual Review of Applied Linguistics* 20: 213–23. [CrossRef]

Shea, Christine. 2019. Dominance, proficiency, and Spanish heritage speakers' production of English and Spanish vowels. *Studies in Second Language Acquisition* 41: 123–49. [CrossRef]

Simonet, Miquel. 2014. Phonetic consequences of dynamic cross-linguistic interference in proficient bilinguals. *Journal of Phonetics* 43: 26–37. [CrossRef]

Simonet, Miquel, and Mark Amengual. 2020. Increased language co-activation leads to enhanced cross-linguistic phonetic convergence. *International Journal of Bilingualism* 24: 208–21. [CrossRef]

Simonet, Miquel, Marcos Rohena-Madrazo, and Mercedes Paz. 2008. Preliminary evidence for incomplete neutralization of coda liquids in Puerto Rican Spanish. In *Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*. Edited by Laura Colantoni and Jeffrey Steele. Somerville: Cascadilla Proceedings Project, pp. 72–86.

Solis-Barroso, Cecilia, and Sara Stefanich. 2019. Measuring language dominance in early Spanish/English bilinguals. *Languages* 4: 62. [CrossRef]

Stefanich, Sara, and Jennifer Cabrelli. 2016. L2 Acquisition of the Spanish palatal nasal. Paper presented at New Sounds 2016: 8th International Conference on Second Language Speech, Aarhus, Denmark, June 10–12.

Trofimovich, Pavel, and Wendy Baker. 2006. Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of l2 speech. *Studies in Second Language Acquisition* 28: 1–30. [CrossRef]

van Leussen, Jan Willem, and Paola Escudero. 2015. Learning to perceive and recognize a second language: The L2LP model revised. *Frontiers in Psychology* 6: 1000. [CrossRef] [PubMed]

Yao, Yao, and Charles Chang. 2016. On the cognitive basis of contact-induced sound change: Vowel merger reversal in Shanghainese. *Language* 92: 433–67. [CrossRef]

languages

Article

Redefining Sociophonetic Competence: Mapping COG Differences in Phrase-Final Fricative Epithesis in L1 and L2 Speakers of French

Amanda Dalola [1,*] and Keiko Bridwell [2,*]

1 Department of Languages, Literatures & Cultures, University of South Carolina, Columbia, SC 29208, USA
2 Department of Linguistics, University of Georgia, Athens, GA 30605, USA
* Correspondence: dalola@mailbox.sc.edu (A.D.); keiko.bridwell@uga.edu (K.B.)

Received: 9 September 2020; Accepted: 2 November 2020; Published: 12 November 2020

Abstract: This article presents a study of measures of center of gravity (COG) in phrase-final fricative epithesis (PFFE) produced by L1 and L2 speakers of Continental French (CF). Participants completed a reading task targeting 98 tokens of /i,y,u/ in phrase-final position. COG measures were taken at the 25%, 50% and 75% marks, normalized and submitted to a mixed linear regression. Results revealed that L2 speakers showed higher COG values than L1 speakers in low PFFE-to-vowel ratios at the 25%, 50%, and 75% marks. COG measures were then categorized into six profile types on the basis of their frequencies at each timepoint: flat–low, flat–high, rising, falling, rising–falling, and falling–rising. Counts of COG profile were then submitted to multinomial logistic regression. Results revealed that although L1 speakers produced predominantly flat–low profile types at lower percent devoicings, L2 speakers preferred multiple strategies involving higher levels of articulatory energy (rising, falling, rise–fall). These results suggest that while L1 speakers realize PFFE differently with respect to phonological context, L2 speakers rely on its most common allophone, strong frication, in most contexts. As such, the findings of this study argue for an additional phonetic dimension in the construct of L2 sociophonetic competence.

Keywords: sociophonetics; competence; fricative epithesis; vowel devoicing; center of gravity; French; acquisition

1. Introduction

Phrase-final fricative epithesis (PFFE), a phenomenon also known in the literature as phrase-final vowel devoicing (PFVD), refers to a well-attested phenomenon in Continental French (CF) in which breath group-final vowels lose their voicing and produce a short burst of high-frequency aperiodic energy, akin to a fricative, e.g., *mais oui_hhh* [mɛwi**ç**], *merci beaucoup_hhh* [mɛʁsibоku**x**] (see Figure 1). The first linguistic description of this phenomenon described it as the emergence of "sharp, phrase-final whistles" (Fónagy 1989); subsequent research witnessed a split in nomenclature, with North American researchers often opting for a name focusing on voicing loss—"vowel devoicing/*dévoisement vocalique*" (Fagyal and Moisset 1999; Smith 2002, 2003, 2006; Martin 2004) and most European researchers preferring a name focusing on the emergence of the downstream fricative—"fricative epithesis/*épithèse (consonantique) fricative*" (Fagyal 2010; Candea 2012; Candea et al. 2013). Because the present study will focus on characterizing the spectral and durational qualities of the emergent fricative, we the (North American) authors have explicitly chosen to heed the call of our European predecessors in adopting the term "fricative epithesis" for this discussion.

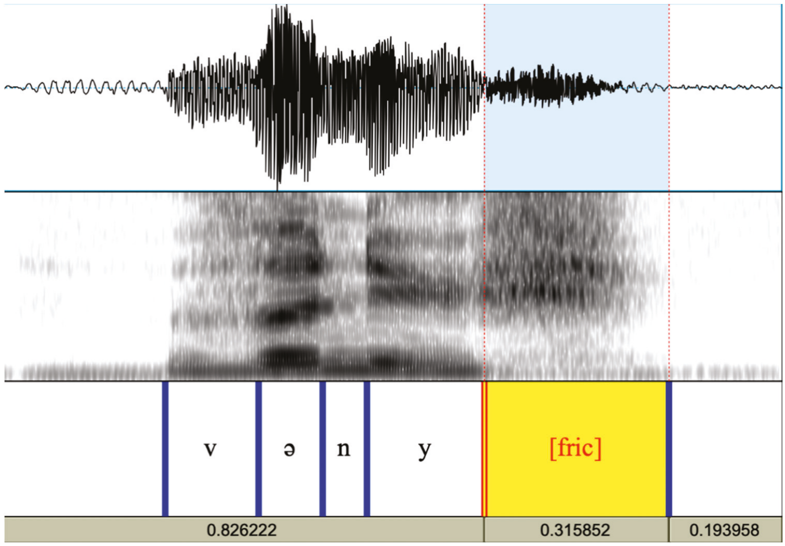


Figure 1. PFFE on the spectrogram: *venu* 'came.' The PFFE corresponds to the final, highlighted segment—characterized by the lack of a voicing band on the spectrogram and aperiodic energy on the waveform—which follows the articulation of the vowel [y], distinguished by its full formant structure on the spectrogram and periodic energy on the waveform.

In the first description of PFFE in the literature, Fónagy hypothesized that not only did its characteristic phrase-final fricatives appear immediately following vowels that had lost a portion of their voicing band, but that he also suspected the fricatives themselves might correspond to the host vowel phonetically in terms of their backness dimension. Citing the *ich-Laut/ach-Laut* harmony phenomenon in standard German, in which the backness value of a voiceless fricative is selected by the backness value of its preceding vowel, he hypothesized that the fricatives epithesized after the high front vowels /i/ and /y/ in French would be more [ç]-like, i.e., front, than those appearing after high back /u/, which would be more [x]-like, i.e., back. This observation was corroborated by Dalola (2015a) who examined measures of center of gravity (COG) (average peak frequency) taken at the 1/4, 1/2 and 3/4 timepoints of PFFE fricatives produced by L1 CF speakers, and found evidence to suggest a three-way distinction in spectral energy at the first two timepoints; however, the spectral differences could not be characterized in terms of sheer [+/– back] and did not persist into the second half of the segment.

1.1. Phonological Predictors of PFFE

The best-studied dimension of PFFE is undoubtedly its phonological distribution. Originally described as occurring in high vowels (Fónagy 1989), subsequent studies documented the occurrence of PFFE in the full inventory of French vowels, including nasals (Smith 2006), but reported the highest rates of PFFE following the high vowels /i,y,u/ (Fagyal and Moisset 1999; Martin 2004; Smith 2003, 2006). When comparing reading passages, role-plays and impromptu conversation, PFFE has been found to occur at significantly higher rates in types of read, i.e., planned, speech (Fagyal and Moisset 1999; Dalola 2014), a finding that is perhaps explained by its higher rates of occurrence at the ends of both the intonation phrase and the declarative phrase (Fagyal and Moisset 1999; Smith 2003), where French sees the arrival of a low tone. Studies have also found an effect for the manner type of the preceding consonant, such that preceding stops condition PFFE at a significantly higher rate than more sonorous manner types, in addition to an effect for lexical frequency, which reports more frequent lexical items as more likely to exhibit the phenomenon than less frequent ones (Dalola 2015b).

1.2. Social Predictors of PFFE

The social distribution of PFFE presents a complex series of macro- and micro-group associations. Early work often described PFFE as occurring in the speech of women (Fónagy 1989; Fagyal and Moisset 1999; Smith 2006); however, later work has reported the variable to be used at similar rates among both men and women (Candea 2012; Candea et al. 2013; Dalola 2014). Fagyal and Moisset (1999), who took a categorical approach to age, found the variable at its highest rates among their youngest (16–35) and oldest (61–85) groups; Dalola (2014), who operationalized age continuously (testing ages 13–83), reported participants as more likely to use PFFE the older they were. From a socioeconomic standpoint, PFFE is often associated with the French middle class (*la bourgeoisie*) (Paternostro 2008; Fagyal 2010). Originally, the variable was associated with Parisians (Fagyal and Moisset 1999; Smith 2006; Fagyal 2010), though in recent years, it has also been documented in the speech of francophones from other metropolitan centers in France, namely Lyon and Strasbourg (Dalola 2014). Further afield, the variable has been described in the speech of French, Belgian and Canadian news anchors (Paternostro 2008; Candea et al. 2013); one study introduced intersectionality into this association by reporting it particularly among young, i.e., inexperienced, news anchors (Candea 2012). Despite the disagreement among social predictors, it is important to pursue research on the characterization of the PFFE variable.

1.3. L2 Speakers and PFFE

Given its salient phonetic energy and robust distribution among native francophone populations, it is somewhat unsurprising to learn that PFFE, despite its status as a sociophonetic variable, is also readily employed by L2 French speakers (Dalola and Bullock 2017). Investigating the nature of L1 and L2 PFFE as produced in different genres of speech, Dalola and Bullock (2017) revealed subtle but nuanced differences at every level of production. For rates of use of PFFE, L1 and L2 speakers performed similarly overall but were motivated by different genres of speech: L1 speakers used more PFFE in role-plays while L2s were more likely to use it when reading wordlists. In terms of duration of PFFE, or the proportional length of the epithesized fricative when compared to its host vowel, larger differences between speaker groups were documented: not only did L1 and L2 speakers produce PFFE segments that were statistically different in length (L1 PFFE length << L2 PFFE length), each group showed sensitivity to a different linguistic parameter: L1s produced longer PFFEs as a reaction to pragmatic shifts (indicated in the prompts to the role-plays), producing longer PFFEs in slower and formal speech, while L2s produced longer PFFEs as a reaction to task shifts, producing longer PFFEs in the wordlist task. Despite the various pragmatic and speaker group effects in this study, no effects were found in the participants for measures of gender or age.

1.4. Perception of PFFE

Differences in L1 versus L2 production of PFFE ushered in a rigorous examination of potential speaker group differences in the variable's perception. Dalola (2016; in progress) reports significant differences in L1 and L2 perceptions of PFFE, namely that L2 speakers perceive it as a positive marker indexing "formality" and "trustworthiness," whereas L1 speakers perceive it variably, sometimes as a positive marker indexing "admirability" and sometimes as a negative marker indexing "emotional affect." Using a matched guide design and exploratory factor analysis, a related form of principal component analysis that partitions out the shared variance of each variable from its unique and error variance to reveal the underlying factor structure (Osborne and Costello 2009), L2 participants in Dalola (2016; in progress) rated users of PFFE similarly for two separate groups of adjectives: *polite, well-educated, speaks clearly, speaks formally* (a category the author refers to collectively as traits of FORMALITY) and *confident, persuasive, I respect X, I trust X* (a category the author refers to collectively as traits of TRUSTWORTHINESS). L1 speakers also rated users of PFFE similarly for two separate groups of adjectives; however, the adjectival members of the groups were both more numerous and compositionally different, with one group including the adjectives *well-educated, professional, speaks clearly, polite, intelligent, patient, confident, persuasive, I trust X, I respect X, I believe what X says, I would like to speak like X* (a category the author refers to collectively as traits of ADMIRABILITY), and the second including the adjectives *aggressive, bourgeois, superficial, bossy, native French speaker, speaks with emotion* (a category the author refers to collectively as EMOTIONAL AFFECT). It should be noted that all the traits that make up the L1 category of FORMALITY are also present in the L2 category of ADMIRABILITY, and that the reason for the difference in category name was due to the author's desire to assign names that applied to the full collection of adjectives. No gender effects were found for the voices being rated, however, there was a significant gender effect among those giving ratings, such that women were more likely to assign higher ratings overall.

1.5. Motivation

This article reports on production differences in spectral tendencies in PFFE among L1 and advanced L2 speakers of Continental French. Since PFFE is a sociophonetic marker in CF (Dalola 2014, 2016), it presents an interesting testing ground for comparing spectral values across native and non-native speakers. While previous work has reported production differences in rate and degree of devoicing between native and non-native French speakers (Dalola and Bullock 2017), it has yet to extend the comparison to investigate the phonetic quality of the variable emergent fricatives. Combined with the many known articulatory differences and false similarities in vowel production between French and English (the L1 of the non-native population in this and previous studies), it is reasonable to expect that articulatory issues may arise, even among advanced L2 speakers (Flege and Hillenbrand 1984; Flege 1985, 1987; among others). The goal of this study is, therefore, to examine and characterize the fricatives epithesized after devoiced vowels using measures of fricative-to-vowel ratio (FVR) (length of fricative divided by length of full vowel) and center of gravity (COG) (average peak frequency reached at designated timepoints during fricative segment). The COG measures will then be used to create a multipoint spectral profile for each fricative that will be classified into more general profile types capturing the overall increase, decrease or static tendency of energy during articulation. We will then use inferential statistics to examine the predictability of each profile type by speaker, fricative-to-vowel ratio and vowel type. After presenting the results unique to each profile type, we will compare the most common profile types produced by each speaker group in order to assess the nature of any significant spectral differences occurring between L1 and advanced L2 PFFE.

1.6. Research Questions and Hypotheses

The current study puts forth the following research questions:

1. Do L1 and advanced L2 French speakers produce PFFE differently in terms of measures of COG?
2. Are speaker group-level differences in the production of PFFE modulated by vowel type and fricative-vowel ratio?

Due to the exploratory nature of the study, predictions will not be offered for each of the research questions, as previous work has not yet diagnosed this aspect of the PFFE variable.

2. Materials and Methods

2.1. Participants

40 speakers of CF participated in the experiment, of which 31 were L1-French and nine L1-English advanced L2-French. All participants were recorded in Paris or Strasbourg in France or in the United States. Among the L1 participants, 23 were women and eight were men, ranging in age from 20 to 66 years (mean = 38.4 years). All L1 speakers were L2 speakers of English, having studied it formally for four or more years and using it in interactions once a week or more. Among the L2 participants, five were women and four were men, ranging in age from 27 to 58 years (mean 38.6 years). L2 speakers

were classified as "advanced" because they had all lived in France for at least two years, had prepared or were preparing an upper-level degree in French and used French regularly in their careers. All L2 speakers were L1 speakers of American English. All subjects gave their informed consent for inclusion before they participated in the study. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the University of South Carolina Institutional Review Board (USC IRB).

2.2. Stimuli

Inspired by the task and pragmatic effects findings of Dalola and Bullock (2017) and several studies' reports of PFFE's robustness among news anchors reading off teleprompters (Paternostro 2008; Candea 2012; Candea et al. 2013), participants were asked to complete a reading task that consisted of 106 single sentences containing 98 phrase-final tokens of /i,y,u/, occurring after all licit (C)C(C) onset sequences in one- to three-syllable real words in French (see Table 1 for a breakdown of consonant environments).

Table 1. Stimuli consonant onset sequences.

Tokens per Sequence	Consonant Onset Sequences
3	/p,t,k,b,d,g,f,s,ʃ,v,z,ʒ,r,l,m,n,kr,tr,bl,gr,fl/
2	/pl,pr,gl,dr,fr,ʒr,ʃr/
1	/ɥ,j,w,kl,bl,pɥ,tɥ,kɥ,bɥ,dɥ,fɥ,sɥ,nɥ,lɥ,lw,bw,fw,lw,nw,plɥ,trɥ/

2.3. Procedure

Participants were presented with sentences one at a time on a MacBook Pro via Microsoft Powerpoint and told to read each one aloud, imagining they were reading a story to a native francophone listener. Participants were instructed to read each sentence twice and to repeat any trials from the beginning in the event of a disfluency. As they read aloud, participants were recorded via a head-mounted unidirectional cardioid microphone (SHURE WH20) plugged into a solid-state digital recorder (Marantz PMD 660) digitized at 44.1 kHz (16 bit). The task was completed under the direction of the L1-English advanced L2-French researcher; it was self-paced and participants were given as much time as they needed to complete it.

2.4. Acoustic Measurements

From the resulting recordings, target vowels were identified, delimited, and labeled manually in Praat based on the spectrographic and time displays, beginning at the onset of voicing and formant structure and ending with the end of formant structure. Instances of PFFE were counted as part of the vowel and were included in the overall duration (measured in milliseconds). Each target vowel was then inspected for the presence of PFFE, which was labeled and measured for duration on a separate tier. A derived measure of the *fricative-vowel ratio (FVR)* was calculated by dividing the length of frication by the length of the full vowel, i.e. the vocalic portion plus frication, as illustrated in Figure 2.

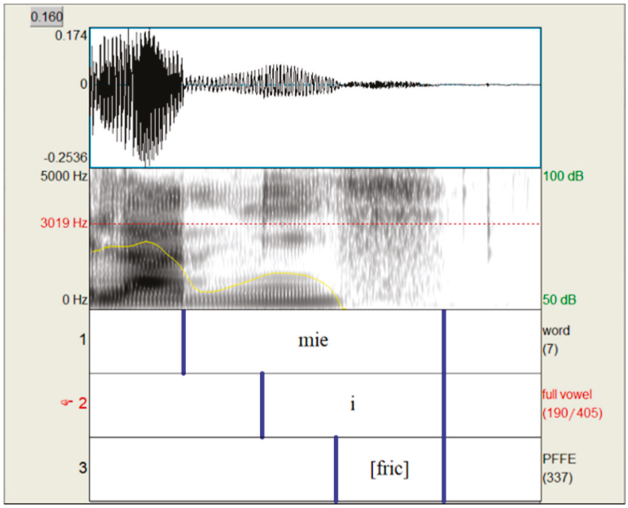


Figure 2. The segmentation of *mie* 'crumb.' With a full-vowel length of 287 ms and a fricative length of 171 ms, the FVR of this token is 59.6%.

A script targeting the devoicing tier divided each instance of PFFE into quartiles, labeling timepoints at the 25%, 50%, and 75% marks; a subsequent script measured center of gravity (COG) at each of these points (Erker 2010).

To control for effects of variation in vocal tract length, COG values were normalized (Shadle and Mair 1996), according to a technique adapted from Toda (2007), as shown in (1):

$$\mathrm{COG}_{\mathrm{norm}} = s_i \times \mathrm{COG}, \tag{1}$$

in which the speaker-dependent coefficient s_i was calculated by (2):

$$s_i = \mathrm{COG}_{\mathrm{avg}}/\mathrm{COG}_i, \tag{2}$$

where COG_i refers to the average COG value of participant i, and $\mathrm{COG}_{\mathrm{avg}}$ refers to the average COG value across all participants. Henceforth, COG will be used to refer to this normalized center of gravity variable, and Hz to the normalized units used to quantify COG.

3. Results

3.1. COG Statistical Treatment

Out of 7942 tokens, participants produced a total of 4995 instances of vowels exhibiting PFFE, which formed the corpus for subsequent analysis.

Statistical analyses of COG were conducted in the statistical tool R (R Core Team 2017). Using *lmer()* from the package *lmerTest* (Kuznetsova et al. 2017), a mixed-effects linear regression model was performed for each timepoint, with *COG* as the dependent variable; *vowel*, *speaker group*, and *FVR* as independent variables; and *participant* treated as a random effect. Visualizations were generated using effects data from the package *effects* (Fox 2003; Fox and Weisberg 2019).

3.2. COG Results

Timepoint 1 (25%). The full model for COG of PFFE at Timepoint 1 is pictured in Table 2. There was an interaction effect between *vowel* and *FVR*, such that higher PFFE-to-vowel ratios corresponded to higher COG values, with the identity of the vowel strongly affecting the rate of increase. As shown in Figure 3, the three vowels showed similar COG values as FVR approached 0%, but exhibited strong differences as FVR increased, with /i/ having a stronger rate of increase than /y/ or /u/. /y/ also showed slightly higher COG values than /u/ across all FVRs.

Table 2. Mixed-effects linear regression model for COG at Timepoint 1.

	Estimate (Lower, Upper CI)	*p*-values	
(Intercept)	536.48 (290.07, 799.00)	0.000	****
Vowel /i/	−402.40 (−662.95, −142.67)	0.003	**
Vowel /u/	−424.28 (−737.23, −113.54)	0.008	**
Speaker group L2 French	471.38 (77.59, 867.21)	0.021	*
FVR	1976.94 (1626.89, 2311.18)	0.000	****
Vowel /i/*Speaker group L2 French	75.68 (−107.00, 258.11)	0.417	
Vowel /u/*Speaker group L2 French	−66.69 (−291.03, 155.82)	0.559	
Vowel /i/*FVR	2014.23 (1630.57, 2398.79)	0.000	****
Vowel /u/*FVR	−366.89 (−845.66, 113.37)	0.134	
Speaker group L2 French*FVR	−624.81 (−1111.46, −151.22)	0.011	*
ICC	0.203		

p-values: $p < 0.1$(.), $p < 0.05$ *, $p < 0.01$ **, $p < 0.001$ ***, $p < 0.0001$ ****.

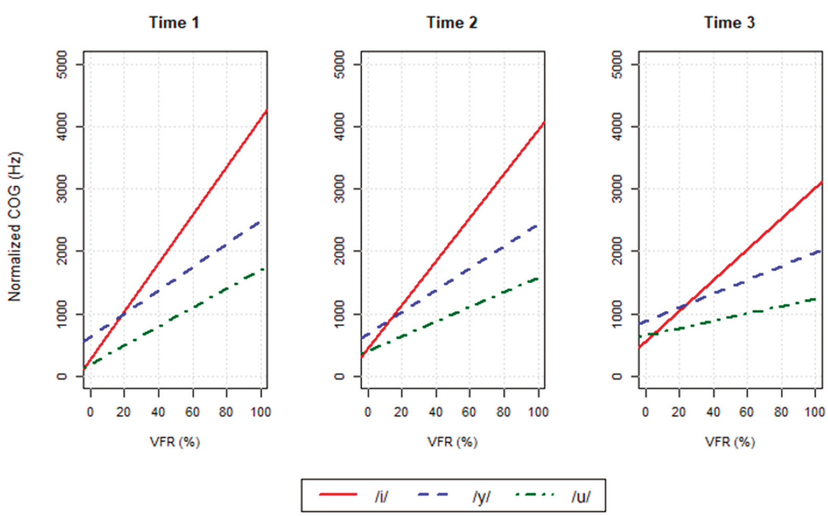


Figure 3. COG by vowel and FVR at all timepoints.

An interaction was also observed between *speaker group* and *FVR*, as visualized in Figure 4, such that L2 speakers exhibited higher COG values than L1 speakers at all FVRs, but to a greater extent at lower percentages.

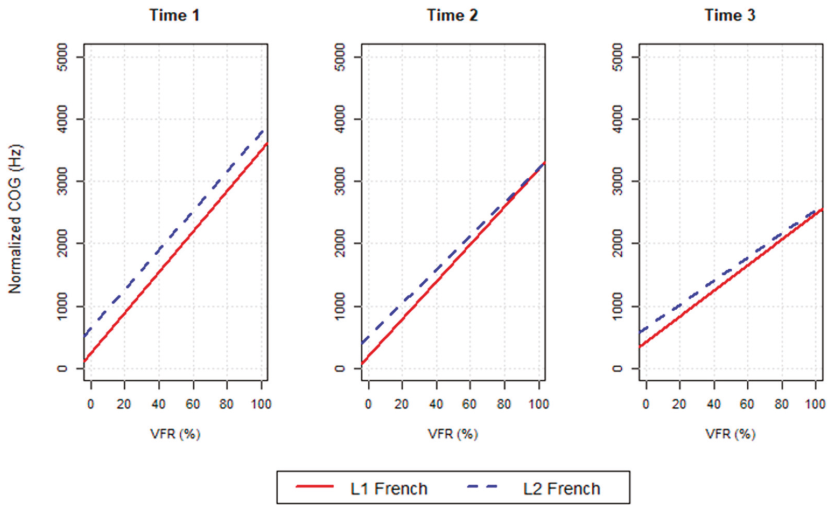


Figure 4. COG by speaker group and FVR at all timepoints.

No significant interaction was observed between *vowel* and *speaker group*. However, there was a main effect for *speaker group*, such that L2 speakers produced PFFE 547.06 Hz higher than L1 speakers. There was also a main effect for *vowel*, such that after all interactions were taken into account, the COG values of /y/ were 402.4 Hz and 424.3 Hz higher than the COG values of /i/ and /u/, respectively (although with respect to the raw data, COG values of /i/ were significantly higher than those of /y/ and /u/). The intraclass correlation coefficient (ICC) of 0.23 suggests low levels of similarity between measurements in the same group, indicating high variability.

Timepoint 2 (50%). A second mixed-effects linear regression model was fit for COG at Timepoint 2, with similar results, as pictured in Table 3. As at Timepoint 1, there was an interaction effect between *vowel* and *FVR*, shown in Figure 3, such that higher PFFE-to-vowel ratios corresponded to higher COG values, with more pronounced differences between COG values at 0% and 100% for /i/ than /y/ or /u/.

Table 3. Mixed-effects linear regression model for COG at Timepoint 2.

	Estimate (Lower, Upper CI)	*p*-values	
(Intercept)	563.21 (325.87, 807.26)	0.000	****
Vowel /i/	−253.95 (−510.67, 1.89)	0.052	.
Vowel /u/	−237.01 (−545.61, 69.01)	0.131	
Speaker group L2 French	578.10 (207.07, 951.79)	0.003	**
FVR	1909.12 (1561.90, 2237.86)	0.000	****
Vowel /i/*Speaker group L2 French	41.21 (−138.82, 220.98)	0.654	
Vowel /u/*Speaker group L2 French	−108.26 (−329.17, 110.93)	0.335	
Vowel /i/*FVR	1745.19 (1367.24, 2124.07)	0.000	****
Vowel /u/*FVR	−602.36 (−1074.00, −128.95)	0.013	*
Speaker group L2 French*FVR	−816.90 (−1289.63, −356.14)	0.001	***
ICC	0.184		

p-values: $p < 0.1$(.), $p < 0.05$ *, $p < 0.01$ **, $p < 0.001$ ***, $p < 0.0001$ ****.

There was also a similar interaction between *speaker group* and *FVR*, such that L2 speakers exhibited higher COG values than L1 speakers at low FVRs (see Figure 4).

No significant interaction was observed between *vowel* and *speaker group*. However, there was a main effect for *speaker group*, such that L2 speakers produced PFFE 619.32 Hz higher than L1 speakers. There was also a trending main effect for *vowel*, such that after all interactions were taken into account, the PFFE of /y/ was produced 253.95 Hz higher than the PFFE of /i/. The ICC measure of 0.184 suggests low levels of similarity between measurements in the same group, indicating high variability.

Timepoint 3 (75%). A final model was fit for COG at Timepoint 3, with similar results to those at Timepoint 1 and Timepoint 2, as pictured in Table 4. As at the previous two time points, there was an interaction effect between *vowel* and *FVR* (see Figure 3), such that higher PFFE-to-vowel ratios corresponded to higher COG values, with more pronounced differences between COG values at 0% and 100% for /i/ than /y/ or /u/.

Table 4. Mixed-effects linear regression model for COG at Timepoint 3.

	Estimate (Lower, Upper CI)	*p*-values	
(Intercept)	779.80 (573.07, 996.62)	0.000	****
Vowel /i/	−331.13 (−565.46, −97.37)	0.006	**
Vowel /u/	−171.56 (−453.86, 107.97)	0.231	
Speaker group L2 French	489.36 (185.02, 795.53)	0.002	**
FVR	1381.23 (884.85, 1503.74)	0.000	****
Vowel /i/*Speaker group L2 French	−21.91 (−186.36, 142.61)	0.794	
Vowel /u/*Speaker group L2 French	−200.51 (−402.05, −0.38)	0.051	.
Vowel /i/*FVR	1381.23 (1035.80, 1726.98)	0.000	****
Vowel /u/*FVR	−534.54 (−965.73, −101.90)	0.015	*
Speaker group L2 French*FVR	−573.72 (−980.56, −172.09)	0.006	**
ICC	0.133		

p-values: $p < 0.1$(.), $p < 0.05$ *, $p < 0.01$ **, $p < 0.001$ ***, $p < 0.0001$ ****.

There was also a similar interaction between *speaker group* and *FVR*, such that L2 speakers exhibited higher COG values than L1 speakers at low FVRs (see Figure 4).

A trending interaction was observed between *vowel* and *speaker group* for /u/, such that L1 speakers produced higher PFFE than L2 speakers. A main effect for *speaker group* was also present, such that L2 speakers produced PFFE 467.45 Hz higher than L1 speakers. Finally, there was a main effect for *vowel*, such that after all interactions were taken into account, the PFFE of /y/ was produced 331.13 Hz higher than the PFFE of /i/. The ICC measure of 0.133 suggests low levels of similarity between measurements in the same group, indicating high variability.

All timepoints. As shown in Figure 3, COG values exhibited a tendency to "level out" over time. At high FVRs, the average COG decreased from Timepoint 1 to Timepoint 3; at low FVRs, the average COG increased slightly over time.

3.3. Profile Creation

While COG values appeared to decrease over time, particularly when PFFE made up a larger proportion of the vowel, this observation was based on aggregate data, not the progression of fricative quality within individual tokens. To investigate this more granularly, a variable combining the three timepoints into a single contour was developed, which will subsequently be referred to as *profile*.

The variable of *profile* was operationalized according to the following procedure. Normalized COG values were first categorized by binning the data: since 98.7% of the data fell under 6000 Hz, the range of frequencies was equally split into levels *Low* (0–2000 Hz), *Medium* (2000–4000 Hz), and *High* (4000+ Hz). For each token, the three timepoints were then combined to form a three-letter profile designation describing the COG pitch over the course of the frication (e.g., HML for a token progressing from high, to medium, to low).

The resulting 27 designations (LLL, LLM, LLH, etc.) were subsequently divided into profile types based on the overall shape which they represented: *flat*, *rising*, *falling*, *rise–fall*, and *fall–rise*. Since a different type of PFFE (vowel devoicing producing only voicing loss, as opposed to vowel devoicing producing only fricative epithesis) appeared to be represented by LLL, LLL was separated from the other members of the "flat" level and labeled as "flat–low." The resulting six levels were used in all subsequent analyses. The correspondence between letter designations and profile types is shown below, in Table 5.

Table 5. Profile categorization.

Profile Type	Categorical COG Designations	Token Count	Token %
Flat-low	LLL	2159	43.2%
Flat-high	MMM, HHH	840	16.%
Rising	LLM, LLH, LMM, LMH, LHH, MMH, MHH	438	8.8%
Falling	MLL, MML, HLL, HML, HMM, HHL, HHM	1051	21.0%
Rise-fall	LML, LHL, LHM, MHL, MHM	384	7.7%
Fall-rise	MLM, MLH, HLM, HLH, HMH	123	2.5%

3.4. Profile Statistical Treatment

All statistical analyses of *profile* were conducted in R (R Core Team 2017). Chi-square tests of the relationship between *profile* and each of the independent variables *vowel*, *speaker group*, and *FVR* were performed using *chisq.test()*. Using *multinom()* from the package *nnet* (Venables and Ripley 2002), a multinomial logistic regression model was also performed, with *COG* as the dependent variable and *vowel*, *speaker group*, and *FVR* as independent variables. Visualizations were generated using effects data from the package *effects* (Fox and Hong 2009; Fox and Weisberg 2019).

3.5. Profile Results

In order to determine whether the different vowels favored different profile types, a chi-square test was conducted on the variables of *profile* and *vowel*. This revealed a significant effect for vowel on profile type ($\chi^2(10) = 1555.2$, $p < 0.0001$). As shown in Table 6, the PFFE of /i/ was most frequently realized with a high, flat COG production, or a COG production that began at high values and decreased throughout the course of the vowel. /y/ and /u/, on the other hand, were most often realized with a low, flat PFFE production. This was particularly true for /u/, which was categorized as flat–low 84.8% of the time. The results of the chi-square test are visualized via correlation plot in Figure 5.

Table 6. Profile distribution by vowel, expressed in % of each vowel.

	Flat Low	Flat High	Rising	Falling	Rise-Fall	Fall-Rise	Total	*n*
/i/	19.51	29.20	9.86	28.76	9.37	3.30	100.00	2436
/y/	51.71	7.88	10.21	19.55	8.24	2.41	100.00	1371
/u/	84.80	0.53	4.66	5.98	3.34	0.70	100.00	1138
Overall	43.22	16.82	8.77	21.04	7.69	2.46	100.00	4995

To determine whether L1 and L2 speakers of French favored different profile groups, another chi-square test was conducted on *profile* and *speaker group*. This test revealed a significant effect for speaker group ($\chi^2(5) = 29.104$, $p < 0.0001$), such that while both groups used a low, flat frication profile more frequently than any other type, L1 speakers used a flat–low profile more often than L2s, and L2 speakers used a flat–high profile more frequently than L1s (see Table 7). All other profile types appeared to be roughly equal across speaker groups. The results of the chi-square test are visualized in a correlation plot in Figure 6.

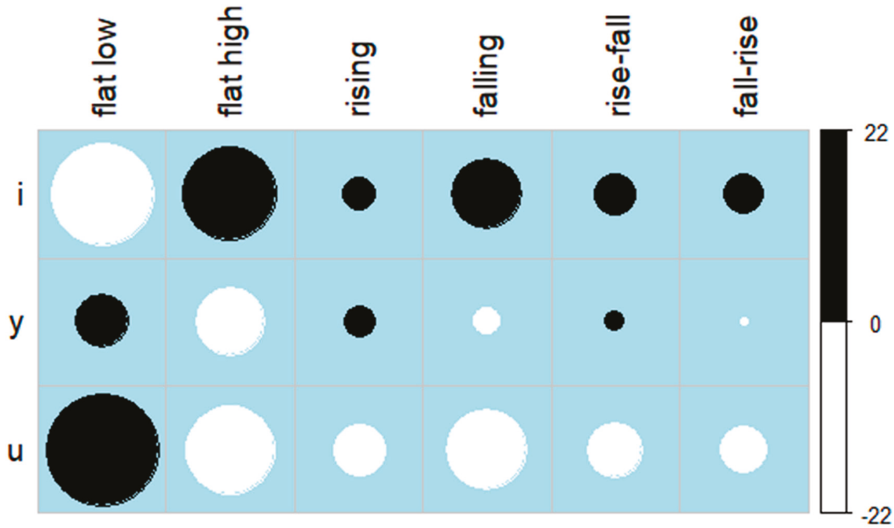


Figure 5. Correlation plot showing residuals of profile type and vowel. Black circles represent positive correlation between variable level pairs, white circles represent negative correlation, and circle radius represents correlation strength.

Table 7. Profile distribution by speaker group, expressed in % of each group.

	Flat Low	Flat High	Rising	Falling	Rise-Fall	Fall-Rise	Total	*n*
L1	44.39	15.39	8.66	20.82	8.23	2.51	100.00	3742
L2	39.74	21.07	9.09	21.71	6.07	2.31	100.00	1253
Overall	43.22	16.82	8.77	21.04	7.69	2.46	100.00	4995

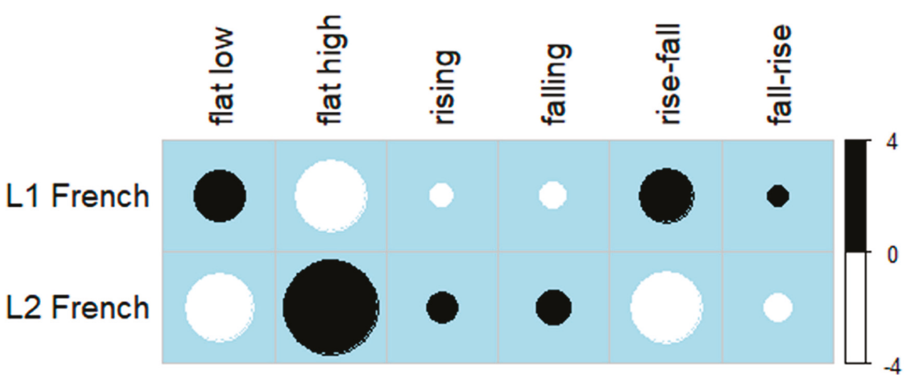


Figure 6. Correlation plot showing residuals of profile type and speaker group.

In order to conduct a chi-square test on *FVR*, its values were binned into five categories with a range of 20%, as shown in Table 8. This test revealed a significant effect for FVR ($\chi^2(20) = 665.92$, $p < 0.0001$). When PFFE made up 60% or less of the vowel, flat–low profiles made up the majority of tokens. It was still the most common profile type at 60–80% percent PFFE, but flat–high and falling were also frequent; at 80–100% PFFE, flat–high was the most common profile type. The results of the chi-square test are visualized in a correlation plot in Figure 7.

Table 8. Profile distribution by FVR, expressed in % of each bin.

	Flat Low	Flat High	Rising	Falling	Rise-Fall	Fall-Rise	Total	*n*
0–20%	52.56	5.13	11.54	17.95	10.26	2.56	100.00	78
20–40%	60.58	3.42	11.66	13.56	8.87	1.90	100.00	789
40–60%	53.16	5.14	11.46	17.07	9.74	3.43	100.00	1283
60–80%	38.20	22.29	6.60	24.57	6.11	2.22	100.00	1848
80–100%	25.28	33.20	6.82	25.78	6.8	2.11	100.00	997
Overall	43.22	16.82	8.77	21.04	7.69	2.46	100.00	4995

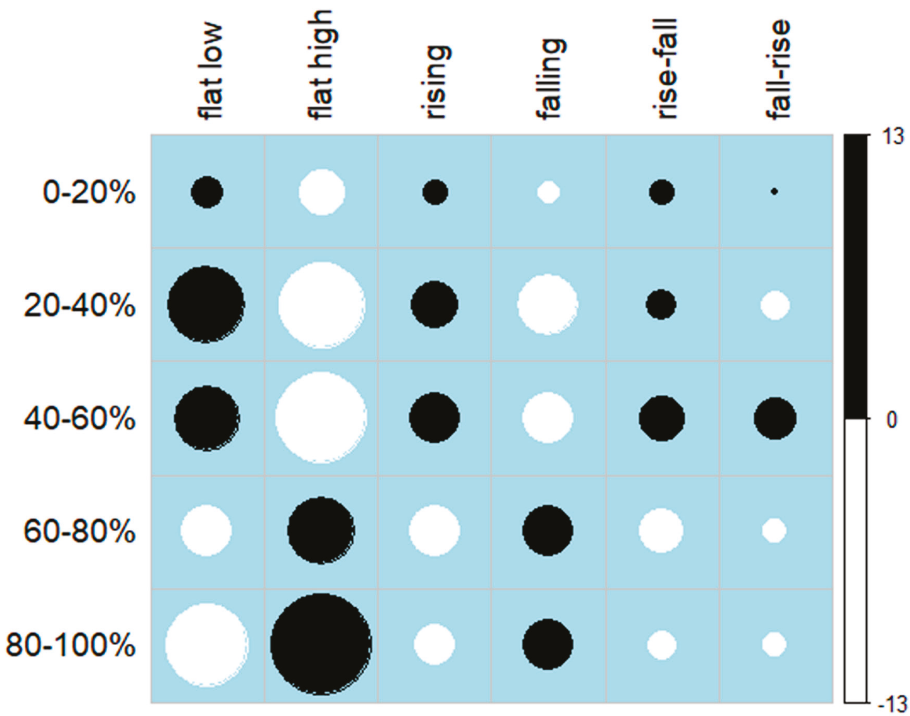


Figure 7. Correlation plot showing residuals of profile type and FVR.

Results from a multinomial logistic regression, shown in Table 9, revealed significant interactions between all three variables. First, an interaction between *vowel* and *speaker group* was present, as shown in Figure 8, such that L1 and L2 speakers showed different distributions of profile type for /i/: L2 speakers showed lower rates of flat–low and higher rates of flat–high and falling, two profile types characterized by high initial energy, relative to L1 speakers. Additionally, L2 speakers utilized more rising tokens than L1 speakers for /y/, indicating that they started off at a low frequency but increased the intensity over the course of PFFE to approximate the high-energy fricative.

Table 9. Multinomial logistic regression model for profile.

	Flat High		Rising		Falling		Rise-Fall		Fall-Rise	
	Est.	***p***	**Est.**	***p***	**Est.**	***p***	**Est.**	***p***	**Est.**	***p***
(Intercept)	−5.772	00.000	−2.533	0.000	−2.968	0.000	−2.665	0.000	−4.504	0.000
/y/	1.489	00.003	0.502	0.195	0.499	0.139	0.363	0.384	0.741	0.303
/u/	−0.264	0.863	0.436	0.399	−0.242	0.618	−0.532	0.388	−1.332	0.339
L2 French	2.700	0.000	−2.802	0.000	1.527	0.000	1.630	0.000	1.972	0.004
FVR	9.436	0.000	1.677	0.000	5.559	0.000	3.488	0.000	4.727	0.000
/y/*L2 French	−1.078	0.000	−0.130	0.657	−1.153	0.000	−0.555	0.095	−0.315	0.526
/u/*L2 French	−18.534	0.000	−1.150	0.006	−1.116	0.001	−1.627	0.006	−11.374	0.932
/y/*FVR	−5.869	0.000	−2.802	0.000	−3.184	0.000	−2.653	0.000	−3.724	0.001
/u/*FVR	−7.423	0.001	−4.592	0.000	−4.592	0.000	−3.233	0.001	−2.585	0.205
L2 French*FVR	−2.322	0.000	−1.935	0.002	−10.030	0.052	−2.118	0.002	−2.280	0.035

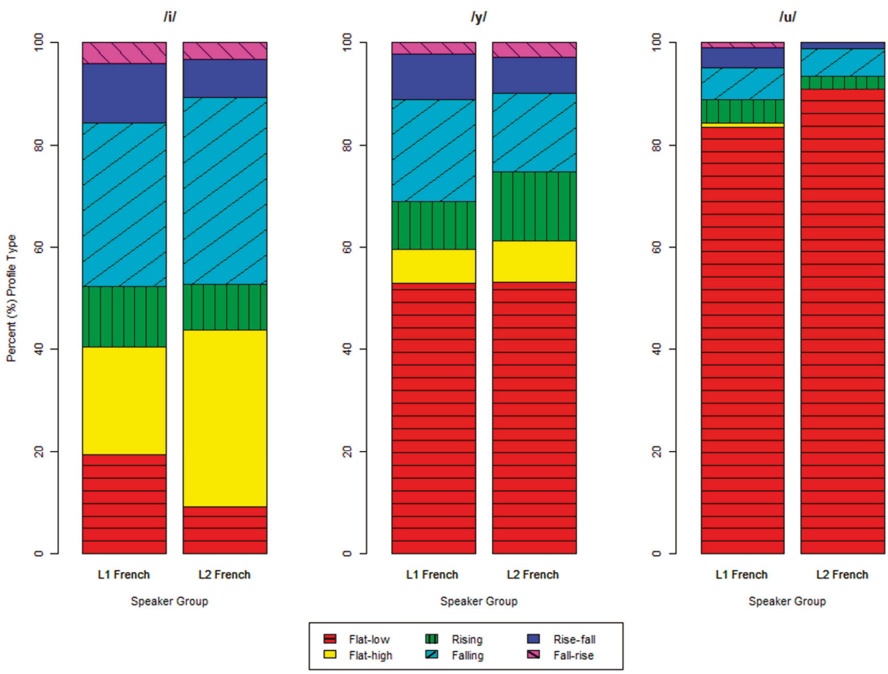


Figure 8. Profile types by vowel and speaker group.

An interaction between *vowel* and *FVR* was also present, such that as FVR increased, the distribution of profile types changed for /i/ and /y/, as shown in the conditional density plot in Figure 9. Starting at approximately 50% FVR (indicated in the figure by a dotted line), flat–low tokens greatly decreased and flat–high tokens greatly increased for /i/, and the proportion of flat–high and falling tokens slightly increased for /y/.

Finally, an interaction was present between *speaker group* and *FVR*, as shown in Figure 10. For FVRs ranging from approximately 40 to 100%, L1 and L2 speakers showed similar profile type distributions. From 0 to 40%, however, L1 speakers predominantly used a flat–low profile, while L2 speakers exhibited much greater variation.

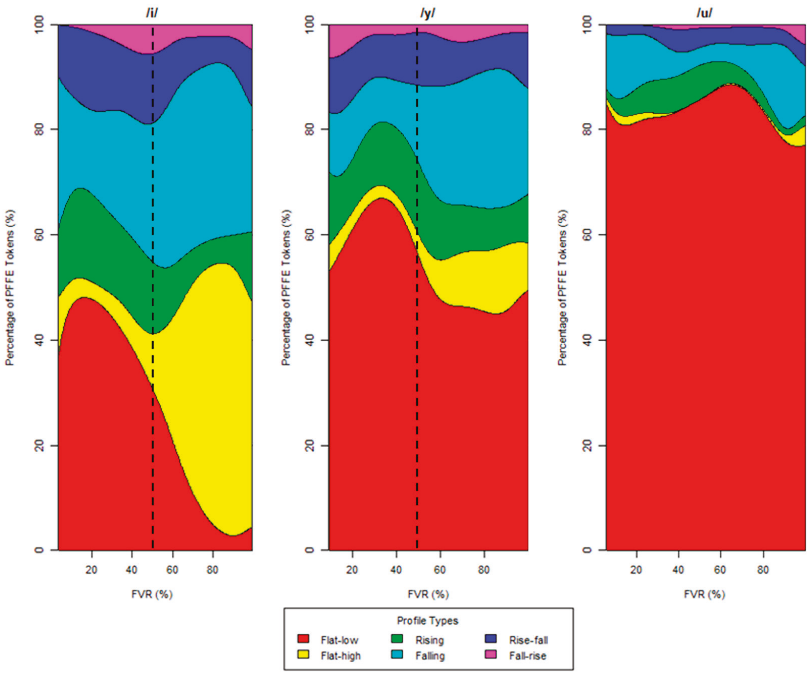


Figure 9. Profile types by vowel and FVR. Each plot shows the proportion of PFFE tokens belonging to each of the profile types at that combination of vowel and FVR. The dotted line at 50% indicates the approximate point at which the distributions of profile types for /i/ and /y/ shifted.

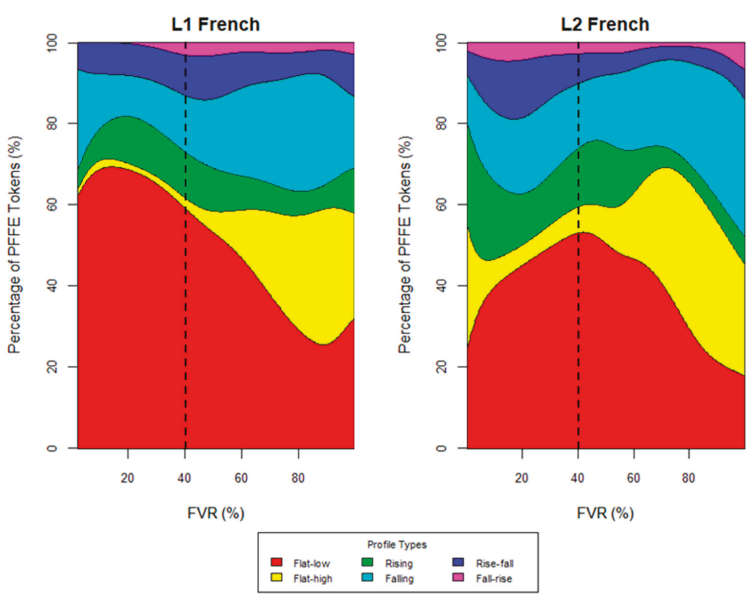


Figure 10. Profile types by speaker group and FVR. The dotted line at 40% indicates the approximate point at which the distribution of profile types shifted.

4. Discussion

4.1. Vowel Findings

The interaction between *vowel* and *FVR* revealed that all high vowels in French do not exhibit PFFE uniformly in terms of its proportion relative to the length of the host vowel. Specifically, the spectral energy in PFFE occurring after vowels /i/ and /y/ showed variable acoustic behavior when the vowel was devoiced to a degree of 50% or more: for /i/, flat–low tokens markedly decreased and flat–high tokens markedly increased, while for /y/, flat–high and falling tokens slightly increased, although flat–low continued to account for a high proportion of tokens. This suggests several acoustic tendencies for PFFE in French. Firstly, the stark decrease in flat–low tokens and increase in flat–high tokens for PFFE in /i/ is indicative of higher-frequency, i.e., more salient, energy being associated with PFFE after /i/ with FVRs between 50 and 100%. When compared to the other lower-frequency, i.e., less salient, COG profiles of other vowels in the high vowel series, the height at which this phenomenon is phonologically most robust, it becomes clear that /i/ devoiced in such a way and to such an extent may be the canonical realization of the PFFE sociophonetic variable for L2 speakers (Dalola and Bridwell 2019). This an important assumption to make when considering L2 speakers' production and distribution of the variable throughout because it may serve as a sort of underlying representation of the phenomenon, which L2 speakers use as a default following other more marked/phonetically dissimilar vowels across French and English where articulatory differences may arise (Delattre 1964). Additionally, the slight increase in flat–high and falling tokens in PFFE in /y/ with FVRs between 50% and 100% is also indicative of a preference for higher-frequency, i.e., more salient, energy, both throughout the segment, as in the case of flat–high, and at the onset of the segment, in the case of falling, but is less notable given that the overall increase is smaller.

4.2. Speaker Group Findings

The interaction between *vowel* and *speaker group* revealed that L1 and L2 French speakers do not realize PFFE uniformly in terms of vowel type. Specifically, the spectral energy in PFFE occurring after vowels /i/ and /y/ showed variable acoustic behavior across speaker group: for /i/, L2 speakers used lower rates of flat–low and higher rates of flat–high and falling, two profile types characterized by high initial energy when compared to L1 realizations, while for /y/, L2 speakers employed more rising tokens than L1s, indicating a build in intensity throughout the course of PFFE to approximate a higher-energy fricative. Both of these findings, when considered in concert with the vowel findings from this study, seem to suggest that L2 speakers may be hyperarticulating PFFE after vowels /i/ and /y/, perhaps as a reflection of how they how understand the phenomenon to sound as a stand-alone segment, i.e., as it does in its most salient form—after vowel /i/ with an FVR between 50 and 100%. Given the relatively lower-energy profiles found in L1 PFFE occurring after these same vowels, the patterning of this phenomenon across speaker groups can be nicely accounted for via Lindblom (1990) H & H Theory, in which speakers vary articulatory clarity according to the information needs of their listener. In this account, the L1 speakers, influenced by phonological context and without a need to communicate sociophonetic information to their unknown, imagined L1 listener (Bell 1984), manifest a sort of hypospeech that is underarticulated and focused on rendering the speech just intelligible enough to be recognized, while L2 speakers, wanting to accurately realize the text but also wishing to signal to their L1 listener their awareness of PFFE as a sociophonetic marker of polished French, manifest a sort of hyperspeech that not only renders optimally intelligible phonemic articulations but overemphasizes certain phonetic features at the expense of maximum articulatory effort. This theory is supported by the previous work examining L1 versus L2 perceptions of the variable, in which it was found that L2 speakers construe PFFE as being associated with features of TRUSTWORTHINESS and FORMALITY (Dalola 2016; in progress), the second of which has notable social capital for advanced L2 speakers using the L2 in their daily and/or professional life. Similar sociophonetic behavior has been found in white Southern Americans using hyperarticulated [hw] to index educatedness (Bridwell 2019), a phonetic behavior characterized by increased duration of the fricative portion of the segment.

The interaction between *FVR* and *speaker group* revealed that L1 and L2 French speakers do not exhibit PFFE uniformly in terms of its proportion relative to the length of the host vowel. Specifically, the spectral energy in PFFE occurring at different FVRs showed variable acoustic behavior across speaker group: for FVRs ranging from approximately 40 to 100%, L1 and L2 speakers showed similar profile type distributions; for FVRs ranging from 0 to 40%, however, L1 speakers predominantly used a flat–low profile, while L2 speakers exhibited much greater, i.e., higher frequency, more salient, variation. This finding observes another instance of L1 speakers favoring low-energy profiles in the articulation of PFFE, while L2 speakers exhibit more and different higher-energy profiles. This is not only consonant with the H & H Theory (Lindblom 1990) offered above but also highlights the greater degree of variation among the L2 population, which may be suggestive of varied exposure times and degrees of involvement in L1 communities where the use of sociophonetic variables, including PFFE, are robust (Dalola and Bullock 2017).

4.3. Implications for Sociophonetic Competence

In light of the findings from this study, we are now able to add an additional parameter to the definition of "sociophonetic competence," as laid out by Dalola and Bullock (2017). Previous work on the sociophonetic variable of PFFE in CF has demonstrated that it is not enough for L2 speakers to have awareness of a sociophonetic variable in their L2 for them to use it at similar rates or durations as their L1 counterparts, or even in the same types of pragmatic and phonological contexts. This study has instead identified an additional dimension of L2 mastery, namely that of phonetic quality of use. Such a mastery at the level of production would also imply a heightened sensitivity to the perception of these sound variations, affording speakers the ability to decode an additional layer of meaning in an L2. Previous sociophonetic work examining COG has demonstrated its ability to work in concert with other parameters to index information about the speaker. Zimman (2017) found COG measures to be a marker of masculine voices when considered alongside f0, while Dalola and Bridwell (forthcoming) found COG measures in conjunction with measures of intensity (loudness), to be a marker of L1- or L2-French speaker status. Taken together, it is possible that COG values in French PFFE may not only be indicative of speaker group status, but also constructs of gender.

Whereas the acoustic energy of PFFE realizations seems to vary allophonically for L1 French speakers, that is to say, as predicted purely by phonological context and ease of articulation, it seems to vary sociophonetically and pragmatically for L2 speakers, that is to say, as conditioned by the desire for speakers to signal their sociophonetic awareness to native listeners at structurally and pragmatically acceptable moments.

4.4. Future Directions

Future studies will sample advanced L2 French populations more robustly and subdivide their level of advancedness via quantitative measures, i.e., the Bilingual Language Profile (Birdsong et al. 2012). This study originally benefited from a more balanced sample across speaker groups, but we later elected to filter out individuals from the L2 category because they did not meet our most stringent criteria (they had not lived abroad in a francophone country for a period of 2 or more years). In restricting the L2 group to the most "advanced" of our sample, we hoped to get the clearest picture of whether or not there were any speaker group differences, however, in doing so ended up diminishing some of our predictive power. In addition, we propose that the current findings be tested via a series of perceptual studies that investigate the pragmatic values of PFFE with differing COG measures in both L1- and L2-French populations. In that way, we can isolate what phonetic components of PFFE contribute reliably to perceptual differences and which ones represent mere physiological variation. A subsequent analysis should compare the COG-motivated perceptual differences that exist between L1- and L2-French speakers in order to test the production findings presented in this analysis. Additionally, since COG (a common spectral moment used to diagnose fricatives) was found to be a meaningful descriptor and predictor for fricatives epithesized after phrase-final vowels, it stands to reason that the other spectral moments (standard deviation, skewness and kurtosis) may also be relevant metrics in characterizing the PFFE variable.

5. Conclusions

The present study has investigated spectral production differences in PFFE as produced by L1 and L2 speakers of French. It has suggested that, even at advanced levels of proficiency and similar rates of use, L2 users do not necessarily realize or distribute the subphonemic properties of sociophonetic variables in nativelike or consistent ways. Future research would do well to query the number and nature of these socially-conditioned subphonemic variables even more (Dalola and Bridwell, forthcoming; Dalola, forthcoming), with the goal of ultimately testing the percepts of their variable forms among both speaker populations.

Author Contributions: Conceptualization, A.D.; methodology, A.D.; software, K.B.; validation, A.D. and K.B.; formal analysis, A.D. and K.B.; investigation, A.D.; resources, A.D. and K.B.; data curation, K.B.; writing—original draft preparation, A.D. and K.B.; writing—review and editing, A.D. and K.B.; visualization, K.B.; supervision, A.D.; project administration, A.D.; funding acquisition, NA. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

Bell, Allan. 1984. Language style as audience design. *Language in Society* 13: 145–204. [CrossRef]

Birdsong, David, Libby M. Gertken, and Mark Amengual. 2012. *Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism*. Austin: COERLL, University of Texas at Austin.

Bridwell, Keiko. 2019. The Distribution of [ʍ]: An Acoustic Analysis of Sociophonetic Factors Governing the Wine-Whine Merger in Southern American English. Master's thesis, University of South Carolina, Columbia, SC, USA.

Candea, Maria. 2012. Au journal de RFI-chhh et dans d'autres émissions radiodiffusée-chhhs. Les épithèses consonantiques fricatives. *Le Discours et la Langue* 2: 136–49.

Candea, Maria, Martine Adda-Decker, and Lori Lamel. 2013. Recent evolution of non-standard consonantal variants in French broadcast news. *Interspeech*, 412–16.

Dalola, Amanda. 2014. Un drôle de bruit_hhh: A Sociophonetic Examination of the Production and Perception of Final Vowel Devoicing among L1 and L2 Speakers of French. Ph.D. dissertation, University of Texas, Austin, TX, USA.

Dalola, Amanda. 2015a. An acoustic profile of phrase-final devoicing in Continental French. Paper presented at LSRL 45, UNICAMP, Campinas, Brazil, May 6–9.

Dalola, Amanda. 2015b. The role of vowel type, preceding consonant and lexical frequency on final vowel devoicing in Continental French. Paper presented at ICPhS18, Glasgow, UK, August 10–14.

Dalola, Amanda. 2016. When L1 & L2 French speakers perceive phrase-final vowel devoicing. Paper presented at NWAV 2016, Vancouver, BC, Canada, November 3–6.

Dalola, Amanda, and Keiko Bridwell. 2019. The shape of [u]: Towards a typology of Final Vowel Devoicing in Continental French. Paper presented at ICPhS19, Melbourne, Australia, August 5–9.

Dalola, Amanda, and Keiko Bridwell. forthcoming. Revisiting Sociophonetic Competence: Variable spectral moments in phrase-final fricative epithesis for L1 & L2 speakers of French. Submitted to *Linguistics Symposium on Romance Languages (LSRL) Proceedings*. Under review.

Dalola, Amanda, and Barbara Bullock. 2017. On Sociophonetic Competence: Phrase-final Vowel Devoicing in Native and Advanced L2 Speakers of French. *Studies in Second Language Acquisition* 39: 769–99. [CrossRef]

Dalola, Amanda. forthcoming. Bossy is in the ear of the beholder: Phrase-final fricative epithesis among L1 & L2 speakers of French. To be submitted to *Journal of French Language Studies*. In progress.

Delattre, Pierre. 1964. Comparing the vocalic features of English, German, Spanish, and French. *IRAL* 2: 71–97. [CrossRef]

Erker, Daniel G. 2010. A subsegmental approach to coda /s/ weakening in Dominican Spanish. *International Journal of the Sociology of Language* 203: 9–26. [CrossRef]

Fagyal, Zsuzsanna. 2010. *Accents de banlieue: Aspects prosodiques du français populaire en contact avec les langues de l'immigration*. Paris: l'Harmattan.

Fagyal, Zsuzsanna, and Christine Moisset. 1999. Sound change and articulatory release: Where and why are high vowels devoiced in Parisian French. *ICPhS* 14: 309–12.

Flege, J. E. 1985. Production of /t/ and /u/ by monolinguals and French-English bilinguals. *The Journal of the Acoustical Society of America* 77: S85. [CrossRef]

Flege, James E. 1987. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15: 47–65. [CrossRef]

Flege, James E., and James Hillenbrand. 1984. Limits on phonetic accuracy in foreign language speech production. *The Journal of the Acoustical Society of America* 76: 708–21. [CrossRef]

Fónagy, Ivan. 1989. Le français change de visage? *Revue Romane* 24: 225–54.

Fox, John. 2003. Effect Displays in R for Generalised Linear Models. *Journal of Statistical Software* 8: 1–27. Available online: https://www.jstatsoft.org/article/view/v008i15 (accessed on 12 August 2020). [CrossRef]

Fox, John, and Jangman Hong. 2009. Effect Displays in R for Multinomial and Proportional-Odds Logit Models: Extensions to the effects Package. *Journal of Statistical Software* 32: 1–24. Available online: https://www.jstatsoft.org/article/view/v032i01 (accessed on 12 August 2020). [CrossRef]

Fox, John, and Sanford Weisberg. 2019. *An R Companion to Applied Regression*, 3rd ed. Thousand Oaks: Sage. Available online: https://socialsciences.mcmaster.ca/jfox/Books/Companion/index.html (accessed on 12 August 2020).

Kuznetsova, Alexandra, Per B. Brockhoff, and Rune H. B. Christensen. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82: 1–26. [CrossRef]

Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H&H theory. In *Speech Production and Speech Modelling*. Dordrecht: Springer, pp. 403–39.

Martin, Pierre. 2004. Dévoisement vocalique en français. *La linguistique* 40: 3–22. [CrossRef]

Osborne, Jason W., and Anna B. Costello. 2009. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Pan-Pacific Management Review* 12: 131–46.

Paternostro, Roberto. 2008. Le dévoisement des voyelles finales. *Rassegna italiana di Linguistica applicata* 3: 129–58.

R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. Vienna: R Found. Stat. Comput.

Shadle, Christine H., and Sheila J. Mair. 1996. Quantifying spectral characteristics of fricatives. Paper presented at Fourth International Conference on Spoken Language Processing, ICSLP96, Philadelphia, PA, USA, October 3–6; vol. 3, pp. 1521–24.

Smith, Caroline L. 2002. Prosodic Finality and Sentence Type in French. *Language and Speech* 45: 141–78. [CrossRef] [PubMed]

Smith, Caroline L. 2003. Vowel Devoicing in Contemporary French. *Journal of French Language Studies* 13: 177–94. [CrossRef]

Smith, Caroline L. 2006. *In and Out of the Laboratory: Using a Variety of Data Sources to Study Variation Source in Speech [Poster]*. Barcelona: LabPhon06.

Toda, Martine. 2007. Speaker normalization of fricative noise: Considerations on language-specific contrast. Paper presented at ICPhS16, Saarbrücken, Germany, August 6–10; pp. 825–28.

Venables, William N., and Brian D. Ripley. 2002. *Modern Applied Statistics with S*, 4th ed. New York: Springer.

Zimman, Lal. 2017. Gender as stylistic bricolage: Transmasculine voices and the relationship between fundamental frequency and/s. *Language in Society* 46: 339–70. [CrossRef]

languages

MDPI

Article

A Phonetic Account of Spanish-English Bilinguals' Divergence with Agreement

Laura Colantoni [1,*], Ruth Martínez [1], Natalia Mazzaro [2], Ana T. Pérez-Leroux [1] and Natalia Rinaldi [1]

1 Department of Spanish and Portuguese, University of Toronto, Toronto, ON M5S 1K7, Canada; ruth.martinez@mail.utoronto.ca (R.M.); at.perez.leroux@utoronto.ca (A.T.P.-L.); natalia.rinaldi@mail.utoronto.ca (N.R.)

2 Department of Languages and Linguistics, University of Texas at El Paso, El Paso, TX 79968, USA; nmazzaro@utep.edu

* Correspondence: laura.colantoni@utoronto.ca

Received: 17 September 2020; Accepted: 2 November 2020; Published: 11 November 2020

Abstract: Does bilingual language influence in the domain of phonetics impact the morphosyntactic domain? Spanish gender is encoded by word-final, unstressed vowels (/a e o/), which may diphthongize in word-boundary vowel sequences. English neutralizes unstressed final vowels and separates across-word vocalic sequences. The realization of gender vowels as schwa, due to cross-linguistic influence, may remain undetected if not directly analyzed. To explore the potential over-reporting of gender accuracy, we conducted parallel phonetic and morphosyntactic analyses of read and semi-spontaneous speech produced by 11 Monolingual speakers and 13 Early and 13 Late Spanish-English bilinguals. F1 and F2 values were extracted at five points for all word-final unstressed vowels and vowel sequences. All determiner phrases (DPs) from narratives were coded for morphological and contextual parameters. Early bilinguals exhibited clear patterns of vowel centralization and higher rates of hiatuses than the other groups. However, the morphological analysis yielded very few errors. A follow-up integrated analysis revealed that /a and o/ were realized as centralized vowels, particularly with [+Animate] nouns. We propose that bilinguals' schwa-like realizations can be over-interpreted as target Spanish vowels. Such variable vowel realization may be a factor in the vulnerability to attrition in gender marking in Spanish as a heritage language.

Keywords: Spanish-English bilinguals; gender; vowels; vowel centralization; vowel sequences

1. Introduction

Heritage speakers, i.e., bilinguals who were raised with a home language different from the dominant language of their communities, exhibit various patterns of divergence and cross-linguistic influence. Studies of heritage speakers of Spanish document difficulties with gender agreement and concord for both adults (e.g., Montrul et al. 2008) and children (Gathercole 2002; Montrul and Potowski 2007; Morgan et al. 2013), which may remain at the level of performance (Montrul et al. 2014) or could lead to internal restructuring of the featural system, as proposed by Scontras et al. (2018) and Cuza and Pérez-Tattam (2016). Attrition of the gender system increases across bilingual generations (Martínez-Gibson 2011) and during school years (Cuza and Pérez-Tattam 2016). Given the frequency of the configuration under consideration (a determiner followed by a noun), and the morphological transparency and robustness of the Spanish agreement system, one might speculate that input factors cannot fully account for the vulnerability of this domain beyond lexical errors with low-frequency, gender-opaque nouns. One thing worth keeping in mind is that analyses of bilingual gender consider only some morphosyntactic or lexical factors and leave aside the potential explanatory value of other domains. We turn to the perception-production interface to explore an alternative explanatory source to current accounts of bilingual vulnerability of Spanish gender.

Spanish agreement is primarily encoded by three vowels (/a e o/) appearing in unstressed word-final position, a context where English neutralizes vocalic contrasts. Word-final vowels followed by other vowels in the word-initial position are another locus of the differences between English and Spanish: English has a tendency towards separating these sequences via insertion of glottal stops or pauses (Davidson and Erker 2014), whereas Spanish diphthongizes or fuses vowels (e.g., Aguilar 2010). Is it possible, then, that heritage bilinguals' realization of Spanish vowels in absolute final position and in sequences introduces what Scontras et al. (2015) dubbed "incipient changes" in the input that could eventually trigger changes in gender marking in subsequent learners' cohorts? Is it also possible that linguists working on bilingual morphosyntax may have overlooked differences in the realization of these vowels when coding bilinguals' speech?

Our overarching goal is not to provide a definitive answer, but to probe preliminary data on these two questions. We conduct analyses of read speech and narratives produced by adult early and late bilinguals (defined in terms of onset of exposure prior to vs. after adolescence, respectively). In our approach, we use phonetic analyses to provide a full characterization of how phonetic variability could potentially impact heritage language acquisition, followed by an exploration of how various lexical and morphosyntactic properties associate with gender realization, and finally, we combine phonetic analyses with accuracy ratings. Thus, we will first phonetically analyze word-final unstressed vowels to compare our results with previous studies (see Section 1.2) that found that early bilinguals centralize vowels when compared to late bilinguals and monolinguals. Second, we will analyze the noun phrases extracted from the narratives to determine the distribution of patterns of morphological realizations in bilinguals to test whether early bilinguals differ from late bilinguals and monolinguals in their rate of errors with gender marking and concord (i.e., gender matching to related morphosyntactic categories). Third, and this is the key contribution of our study, we will integrate our phonetic and morphosyntactic results, arguing that the findings cast some doubt on our capacity of reporting agreement accuracy only based on the researcher's auditory transcriptions. Finally, we will use the insights extracted from this dataset to propose a modular interaction hypothesis, which claims that language contact-related variability in one domain (phonetics-phonology) can have important consequences for the acquisition of other domains (grammar). The testing of this latter hypothesis is a long-term goal of our research team and involves the design of an acquisition study that controls for the perception-production of the target vowels and the comprehension and production of agreement and concord. Before we delve into the analysis of our corpus, we pair previous findings from studies on the acquisition of gender with those of phonetic studies on the acquisition of vowels in Spanish-English bilinguals.

1.1. Acquisition of Gender in Spanish-English Bilinguals

Spanish encodes agreement as word-final affixes in the nominal system. Nouns and their dependents (determiners, adjectives, and pronouns) are marked (whether visibly or not) for gender. Gender marking can either have semantic value, as in *niña/niño*, 'girl/boy' or simply be a formal word-marker, as in *pala*, 'shovel' vs. *palo*, 'stick'. Spanish gender is considered transparent as most nouns overtly mark gender by means of the canonical vowel suffix. A few nouns do not directly reflect the gender feature: those ending in consonants, other vowels, or in mismatching vowels.

Monolingual and bilingual children show different patterns of development. Monolingual children learn noun agreement in spontaneous speech by the age of 2;0 (Lopez Ornat 1997; Snyder et al. 2001). Later gender-assignment errors (as identified by the gender of the article; e.g., **la mapa*; 'the-FEM map') reflect the lack of lexical knowledge of the gender of opaque nouns and expresses biases towards phonological gender cues (Pérez-Pereira 1991). Errors in concord (matching gender across the noun phrase) are negligible in the preschool years (Castilla and Pérez-Leroux 2010). Bilingual children reach mastery much later (Barreña 1997; Eichler et al. 2013; Larranãga et al. 2012), and early infant speech in bilinguals does not show the harmonic prenominal vowel patterns used by monolingual infants before they fully acquire determiners (Kuchenbrandt 2005;)e.g., a vaca, '(the-fem cow'). Elicited data allow a direct comparison of attainment of noun-adjective agreement in monolingual and bilingual communities. Bedore and Leonard (2001) tested Mexican American five-year-olds in San Diego. These children, described as having minimal access to English, showed noun-adjective agreement accuracy at 76–91%. In contrast, same-age children tested in Mexico by Grinstead et al. (2008) were at ceiling. Morgan et al. (2013) found that five-year-old US bilinguals displayed few, but significantly more, gender substitution errors than their monolingual counterparts. In contrast, Cuza and Pérez-Tattam (2016) presented school-aged children with morphologically opaque nouns. Bilinguals produced many more errors than same-age monolinguals and were delayed in both their knowledge of gender assignment and of concord. Beyond noting a range of patterns of performance across different bilingual populations, some studies link exposure to Spanish to fewer gender errors (Gathercole 2002). For older children, Montrul and Potowski (2007) found differences between monolinguals and bilinguals, and between simultaneous and sequential bilinguals.

More recent studies (Goebel-Mahrle and Shin 2020), however, failed to find differences between monolingual and bilingual heritage speakers, independently of their age, which ranged between 5 and 11. This was also the case in the word-repetition task included in Montrul et al. (2014) study. Although adult Spanish heritage speakers and L2 learners showed differences from the monolingual comparison group in the other two tasks in the study (a gender monitoring task and a grammaticality judgement task), they did not differ from monolinguals in the word-repetition task. While acknowledging that bilinguals are not a uniform group, we note that the results of the word-repetition task might not necessarily indicate a higher accuracy in production by bilinguals, but in the limitations of auditory transcriptions instead. From a performance point of view, gender is often highly predictable in context. If bilinguals produce a schwa-like vowel rather than the underlying vowel, a transcriber might analyze the item as correct, independently of the quality of the target vowel. A schwa-like vowel insertion allows to maintain the syllable structure of the target word, and introduces acoustic ambiguity. In a predictable context, the transcriber perceives a vowel that cannot be clearly interpreted as a non-target vowel (e.g., *caperucit**a** roj**o***, 'Little Red Riding Hood$_{\text{FEM}}$ red$_{\text{MASC}}$') and transcribes the vowel as [a], in what can be interpreted as a case of *in dubio pro reo*. Why do we think this is the case? Because the literature that we will review below consistently shows that Spanish-English bilinguals exhibit clear patterns of centralization of their vocalic space. We strongly believe that pairing up the phonetic literature on the acquisition of vowels by Spanish-English bilinguals with the findings from the literature on the acquisition of gender by the same group of bilinguals should at least make us wonder if the findings are compatible.

1.2. Acquisition of Vowels in Spanish-English Bilinguals

Spanish vowels (/a e i o u/) vary little in quality and remain contrastive in stressed and unstressed positions (Hualde 2014). Although the literature reports variation in duration (e.g., Delattre 1965) and, to a lesser extent, in quality (Romanelli et al. 2018), Spanish does not show centralization to the extent of English (Navarro Toma's 1970, p. 43). Spanish vowels are learned early, typically by the age of 2;0 (Goldstein and Pollock 2000; Schnitzer and Krasinski 1994). This contrasts with English, where infants begin with a centralized vowel space that expands over time (Gilbert et al. 1997; Kent and Murray 1982; Rvachew et al. 1996). Vowel space differentiation is reported for eighteen-month-olds (Rvachew et al. 2006), but the full English vocalic inventory is mastered later, by 3;0 (Stoel-Gammon and Sosa 2008; Stoel-Gammon and Pollock 2009). This is to be expected, given that English has a large vocalic inventory in the stressed position (Ladefoged 2001). In the unstressed position, the only frequent vowel is schwa [ə] (Rogers 2000). Thus, the unstressed vowel inventory is smaller in English than in Spanish (Hualde 2014).

Studies on the acquisition of Spanish vowels by bilinguals (Ronquest and Rao 2018) report consistent effects of cross-linguistic influence. Bilingual children centralize unstressed vowels

(Gildersleeve-Neumann et al. 2009; Menke 2010),[1] although it is debated whether centralization happens before or after children enter the school system (Gildersleeve-Neumann et al. 2009). Differences in vowel realizations persist into adulthood (Rogers 2012; Ronquest 2016; Willis 2005). Cross-linguistic effects are also reported in perception. Studies of English learners of Spanish suggest late bilinguals tend to confuse high-front with mid-front vowels (/i/ with /e/) and back-mid with low vowels (/o/ with /a/) (Morrison 2003; Morrison 2006). Unlike monolinguals, bilinguals rely on frequency cues rather than on duration (Fox et al. 1994). One perception study on adult heritage Spanish speakers (Mazzaro et al. 2016) found differences, but only in unstressed positions.

These results, however, refer to single vowels. A factor not yet considered is how phonological processes affect word-final vowels, the cross-linguistic differences in this domain, and the potential cross-linguistic interactions. Spanish tends to fuse vowels across words. When both vowels are unstressed, as in *como alfajores* ('I eat cookies'), the highest vowel in the sequence is frequently reduced, ranging from gliding to full deletion. Reduction processes apply to high and non-high vowels equally (Hualde et al. 2008; Vokic and Guitart 2009) across Spanish dialects (Alba 2006; Hutchinson 1974), but reduction is less frequent when one of the vowels is stressed (Colantoni and Hualde 2016; Hualde 2014). Because agreement vowels are in the unstressed word-final position, they are likely modified in running speech. Contrastingly, vowel reduction across word boundaries is rare in English, given that speakers frequently insert glottal stops to separate across-word vocalic sequences (Davidson and Erker 2014). Thus, whereas Spanish prefers diphthongization or deletion, English realizes these sequences as hiatuses.

Crucially, for coarticulation to occur, words must belong to the same intonational phrase or the same prosodic unit. Because nouns and adjectives are prosodified together in Spanish (D'Imperio et al. 2005; Frota et al. 2007), we assume that agreement vowels are often coarticulated in the input. An important question is whether coarticulation affects bilingual patterns of acquisition of suffixal morphology. It has been shown that word-edges play an important role in word recognition (Shoemaker and Rast 2013) and that coarticulation of consonants has a negative effect on lexical retrieval, as speakers struggle to compensate for coarticulated sounds (Mohaghegh 2016). We can thus expect a similar effect for vowels.

1.3. Phonetics and Morphosyntax

Research on the bilingual acquisition of Spanish morphology has explored multiple factors, such as frequency of use, nature of the input, and type of target form, but has not examined the phonetic properties of agreement. This was previously pointed out by Silva-Corvala'n (2014). Phonetic factors, however, are known to be important predictors of functional morphology in the child acquisition literature. Prosody is a common explanation of omission of functional elements, such as articles and clitics (the Prosodic Bootstrapping hypothesis by Lleo' and Demuth 1999; Guasti et al. 2008; Mateu 2015). A robust literature reveals that phonology is key in morphological development. For example, accuracy in production of the English plural -s and third person -s depends on syllable (coda) structure (Ettlinger and Zapf 2011; Song et al. 2009; see also Bernhardt and Stemberger 1998 for an overview). Culberston et al. (2019) show that children use phonological rather than semantic categories when acquiring gender. Phonetic reduction appears to have an impact on the overall course of development, affecting both comprehension and production. Miller and Schmitt (2010) show that the variable realization of the final /s/ in Spanish impacts the acquisition of both plurals and tense agreement (2Sg present). Children growing up in varieties where /s/ is maintained (i.e., not weakened or deleted) acquire these markers earlier, both in comprehension and production, than children growing up in varieties where /s/ is frequently aspirated and deleted. Thus, given evidence that phonetic characteristics of the input affect the acquisition of number (English, Spanish) and person (English),

[1] Kehoe and Lleó (2017) report similar results for German-Spanish bilinguals.

we hypothesize that phonetic variability induced by cross-linguistic influence in the phonetic domain will impact the acquisition of gender in Spanish-English bilinguals.

1.4. Research Questions and Hypotheses

RQ1: What is the phonetic realization of word-final unstressed vowels and across-word vowel + vowel sequences in the three groups?

H1: We hypothesize that early bilinguals (EB) will show a higher rate of vowel overlap (single vowels) than the other two groups. We also hypothesize that monolinguals (M) and late bilinguals (LB) will tend to fuse vowels across words while EBs will tend to separate them.

RQ2: Are there more errors in gender agreement and concord in EBs than in the other groups?

H2: EBs will show a larger proportion of errors than the other two groups.

RQ3: Is vowel centralization being reported as accurate gender marking?

H3: EBs will show a higher rate of centralization when compared with the other two groups, and cases of vowel centralization will be labelled as accurate gender.

RQ4: Do contextual factors (i.e., predictability of the gender of the noun) predict vowel centralization?

H4: Highly predictable DPs are less likely to be fully specified phonetically as they offer redundant information.

2. Materials and Methods

2.1. Participants

A total of thirty-seven participants (*N* = 37) took part in the study: 13 Early Bilinguals, 13 Late bilinguals, whose first language was Spanish, and 11 monolingual speakers serving as the comparison group. All the bilingual participants were residents of El Paso, US who were attending different classes at the University of Texas at El Paso. Participants in the (functionally) monolingual group were Spanish speakers with minimal exposure to and ability in English and were residents of Ciudad Juárez, Mexico. They were either recruited from a beginner-level English for Speakers of Other Languages (ESOL) course, or contacted through social networks.

Following previous research (Montrul 2011; Silva-Corvala'n 2014), the EBs were second- or first-generation immigrants who acquired Spanish during childhood at home or in other natural contexts where a majority language (English) was spoken. They were either born and raised in the US or immigrated permanently to the US at or before the age of 12. The LBs were of first generation Mexican background who arrived in the US after the age of 13 with fully developed L1 grammar. Participants completed an adult language background questionnaire, which elicited information on place of birth, primary language of schooling, patterns of language use, etc. This questionnaire also elicited a self-proficiency judgment in both English and Spanish in the four linguistic skills via a Likert scale, ranging from basic/limited (1) to excellent/native (4). In addition to the self-proficiency measure, participants completed an independent proficiency task, adapted from the *Diploma de Español como Lengua Extranjera* (DELE) (Cuza et al. 2013). Table 1 summarizes Age, Age of Arrival to an English context (AOA), and Length of Residence in an English context (LOR) for all participant groups.

EBs scored an average of 40 in the DELE test, while LBs scored 45. Previous research using this methodology (Cuza et al. 2013; Montrul et al. 2003) considered participants who scored between 40 and 50 points (out of 50) to be 'advanced' learners, those with scores of 30 to 39 were considered to be 'intermediate' learners, and those with scores between 0 and 29 were considered to be 'beginner' learners. In other words, while the proficiency score of the EB group is a bit lower than other groups, all groups have a high overall level of proficiency in Spanish. The EB group included participants born and raised in the US and those who came to the US before the age of six (AOA = 2.5). Their self-proficiency rating in English was near native (3.73/4), while, in Spanish, it was good/fluent (2.9/4). Regarding their patterns of language use, most of the participants (62%) reported using both English and Spanish at home, but they used mainly English at work (75%) and in social situations (46%). Six participants felt more comfortable in English (46%) and six participants (46%) felt equally comfortable in both English and Spanish (46%). Only one participant (8%) selected Spanish as the most comfortable language.

Table 1. Participants' demographic information.

	Early Bilinguals (EB) (n = 13; Female = 7)			Late Bilinguals (LB) (n = 13; Female = 9)			Monolinguals (M) (n = 11; Female = 10)		
Mean age at testing (SD)	23 (2.64)			38 (13.46)			25 (11.13)		
Mean AOA (SD)	2.5 (3.6)			23 (9.37)			20 (4.23)		
Mean LOR (SD)	21 (2.95)			16 (9.6)			08 (0.8)		
Self-Proficiency	English = 3.73/4 Spanish = 2.9/4			English = 2.9/4 Spanish = 3.9/4			English = 1.43/4 Spanish = 4/4		
DELE score	40/50			45/50			45/50		
Language use	SPAN	ENG	BOTH	SPAN	ENG	BOTH	SPAN	ENG	BOTH
Home	15%	23%	62%	77%	8%	15%	100%	0%	0%
School	8%	77%	15%	42%	17%	41%	90%	0%	10%
Work	9%	64%	27%	10%	40%	50%	100%	0%	0%
Social situations	23%	46%	31%	58%	0%	42%	100%	0%	10%
	SPAN	ENG	BOTH	SPAN	ENG	BOTH	SPAN	ENG	BOTH
Most comfortable in	8%	46%	46%	77%	0%	23%	100%	0%	0%

Late bilinguals (*n*= 13) included first generation immigrants from Mexico (mean age at testing = 38; mean AOA = 23; mean LOR = 16). Their self-proficiency rating in Spanish was almost native (3.9/4), and in English it was good/fluent (2.9/4). The proficiency score in the DELE test was 45/50 (advanced proficiency). As for language use, the majority reported speaking more Spanish at home (77%) and in social situations (58%). At school, five participants (N = 5) used both English and Spanish and another five used only Spanish (42%). At work, half of the participants used only English, and 40% used both English and Spanish. When asked which language they felt most comfortable in, the majority (77%) indicated Spanish.

The comparison group consisted of seven recent arrivals in El Paso, Texas, and four residents of Ciudad Juárez, Mexico (mean age at testing = 25; mean AOA = 20; mean LOR = 9 months). Although their AOA of English is earlier than LB (20 vs. 23), these speakers had learned English for a shorter period of time, specifically an average of 9 months. Most of the participants reported speaking more Spanish at home, school, work, and in social situations. They also reported feeling most comfortable speaking Spanish.

All subjects gave their informed consent for inclusion before they participated in the study. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the University of Texas at El Paso IRB.

2.2. Materials and Tasks

To test our hypothesis, we analyzed reading data from "The North Wind and the Sun" and the narrative of the folk tale "Little Red Riding Hood". To elicit the narrative, participants were shown wordless pictures of a children's book based on Perrault's version of the tale. First, participants were shown the pictures as a refresher. When they felt ready, participants recorded the narrative using the images as guidance. All the sessions were conducted in a sound-treated room. Informants were recorded directly onto a laptop computer using Audacity 2.1.2 (available on-line: http://audacityteam.org/) and a Blue Snowball USB microphone. The speech was sampled at 44.1K, and word-final vowels and vocalic sequences (/a e o ae ea oa/) were analyzed with PRAAT (Boersma and Weenink 2001).

2.3. Analysis

The data obtained was subject to two analyses. To analyze the vowels phonetically, we extracted all tokens of /a e o/ in unstressed word-final position as well as all the unstressed across-word vowel sequences (the first vowel in the sequence was one of the three target vowels in the study) using PRAAT. We marked the onset of each vowel or vowel sequence at the beginning of the F1 increase and the offset at the drop in intensity (pre-pausal vowels) or at the beginning of the F1 rise for cases in which the word-final vowel was immediately followed by a word beginning with a consonant. All measurements were taken at zero-crossings. From the reading, we extracted the vowels from nouns, adjectives, and verbs (single vowels: N = 1045; sequences: N = 287) whereas, from the narrative, only vowels that were part of the noun phrase were extracted (single vowels: N = 2954; sequences: N = 1109).[2] Although formant values ware automatically extracted at five points, we will report values at mid-point for single vowels (to minimize the effects of coarticulation with surrounding consonants) and values at five points for sequences. Formant values were subsequently checked manually to inspect values that fell outside the ranges reported in previous studies for Spanish and Spanish-English bilinguals. Data were normalized using the Lobanov (1971) method, as adapted by Nearey (1977) and Adank et al. (2004). Single-vowel results were submitted to two complementary analyses aiming at computing the degree of overlap in vocalic spaces in each group. The first analysis computes the Bhattacharyya affinity scores, following Johnson (2015) and Strelluf (2016). The Bhattacharyya's affinity measures the degree of overlap between two Gaussian distributions (Mak Brian 1996) and is considered to be less sensitive than other approaches when there are imbalances in the sample (Strelluf 2016), which is the case in both tasks but particularly in the narrative. A score of 0 indicates that there is no overlap between vocalic spaces, whereas a score of 1 signals complete overlap. To calculate the Bhattacharyya's affinity score we used the "kernel overlap" function of the {adehabitatHR} R library (Calenge 2006). The second analysis, the computation of convex hulls, as implemented by Haynes and Taylor (2014), allowed us to quantify the percentage of overlap between vowel pairs by calculating the smallest convex shape (polyhedron) that would fit all the given datapoints (Haynes and Taylor 2014, p. 885).[3] Results are reported as percentages and were calculated with the package {phonR} (McCloy 2016).

Vocalic sequences were labeled and then transcribed by one of the authors and then verified by another author. Using the acoustic information available in the spectrogram, we distinguished sequences that were realized together (i.e., no pause or prosodic brake, such as pitch reset between the vowels) from those that were separated by a pause or a glottal stop. We labelled the former as 'diphthong' and the latter as 'hiatus'. This allowed us to clearly separate the sequences that cannot be coarticulated (those labeled as 'hiatus') from those that could (those labeled as 'diphthong'). Thus, the sequences labeled as 'diphthong' may include hiatus realizations, i.e., sequences in which each vowel has frequential and durational values that are similar to those of single vowels (e.g., Aguilar 1999; Borzone de Manrique 1976; Colantoni and Limanni 2010).[4] To determine whether there were frequential differences, sequences that were labeled as diphthongs were acoustically analyzed for their formant trajectories. For the analysis of F1 and F2 trajectories within the unstressed sequences /ae oe ea/,[5] we ran Smoothing Splines ANOVAs (SSANOVAs) using the package {gss} (Gu 2014) to test for statistical differences across the three groups of speakers (EB, LB, M) at five intervals within each sequence. SSANOVAs create smoothing splines for each group by connecting mean data points

---

2 Although no tokens had to be discarded from the reading task, we excluded 72 single vowel tokens from the narrative which were produced with creaky voice.

3 As explained by Haynes and Taylor (2014), these are measurements of overlap and not of statistical significance. As in Haynes and Taylor's study, we are not interested in statistical differences but in the extent to which vocalic spaces overlap, because a greater degree of overlap suggests that the vowels are articulated similarly.

4 Hiatus also differ from diphthongs in the duration and trajectory of formant transitions (e.g., Aguilar 1999).

5 Only these sequences were analyzed because (i) the first vowel was unstressed and (ii) they appeared in the reading task and in the narrative.

through time, as well as 95% Bayesian confidence intervals represented by dotted curves above and below the splines. At the time-points where confidence interval curves do not intersect, two splines are considered significantly different.

To conduct the morpho-syntactic analysis, all noun phrases (NPs) were extracted, representing a total of 2445 analyzed tokens.[6] Determiner phrases (DPs) were coded for configuration (whether it contained a noun, determiner or quantifier, or modifiers, as in (1)), realization of gender and number agreement in each relevant category (nouns, determiners, and adjectives), and overall concord patterns (match/mismatch between constituents as well as target realization for the head noun, as shown in (2)). All nouns were further analyzed for various semantic parameters to explore potential association between form and meaning. These included semantic type of noun in terms of categories pertaining to concreteness and animacy (3) and to individuation and countability (4). Additional coding included whether the noun's initial segment was stressed [a], which, in Spanish singular feminine nouns, leads to an exceptional use of the masculine article (5), as in *el águila* (f) 'the eagle'.[7] Overtly marked gender is indicated in the glosses.

(1) DP configuration

Noun:	lobo	'wolf'
Adjective + Noun	gran lobo	'big wolf-MAS'
Determiner + Noun:	el lobo	'the-MAS wolf-MAS'
Determiner + Covert Noun + Adjective:	el grande	'the-MAS big'
Determiner + Adjective + Noun:	el gran lobo	'the-MAS big wolf-MAS'
Determiner + Noun + Adjective:	el lobo grande	'the-MAS wolf-MAS big'

(2) Agreement: match for agreement features (number or gender) with determiners and/or adjectives, and with target gender of the noun

Correct:	las casas bonitas	'the-FEM pretty-FEM houses-FEM'
Incorrect:	la casa bonito	'the-FEM pretty-MAS house-FEM'

(3) Noun semantic type: semantic features of the noun under analysis

Abstract:	paz	'peace'
Concrete:	coche	'car'
Animate:	gato	'cat'
Human:	chica	'girl'
Event:	fiesta	'party'

(4) Individuation

Mass:	arena	'sand'
Individual:	coche	'car'
Collective:	equipo	'team'
Ambiguous (used as individual):	la policía	'female police officer'
Ambiguous (used as collective):	la policía	'police'
Both (mass or individual):	fruta	'fruit'

All nouns were also coded for morphology and for their morphological relation to other noun lexemes. First, we isolated the final suffix or segment. The first goal was to determine whether gender was visibly expressed or not; that is, whether the noun contained the transparent word markers -a

6 Frozen expressions such as *Fin*, *Colorín Colorado* ('the end') and *lugar* ('place') in the prepositional locution *en lugar de* ('in lieu of') were excluded from the analysis.

7 The morphosyntactic coding was conducted by one of the authors and then was verified by another author. As concerns the error count, this was first checked by one investigator, then re-checked by a second author, and discrepancies (N = 2) were solved by a third author.

(f) and -o (m), a different transparent suffix such as *-ción*, which is uniformly feminine, or whether it contained formally opaque final segments or suffixes (5). We then considered whether the noun entered a gender alternation or not, as in (6), and if so, what was the lexical relationship to the other entry in the alternation (7).

(5) Gender visibility

[nouns with formally visible gender]		
word marker:	niño (m)	'boy'
	niña (f)	'girl'
transparent suffix:	educación (f)	'education'
	cazador (m)	'hunter'
[nouns with no visible gender marking]		
-e:	padre (m)	'father'
-i / -u:	espíritu (m)	'spirit'
consonant:	amor (m)	'love'
inverted (-o for fem. /-a for masc.):	mano (f), foto (m)	'hand, photo'

(6) Gender alternation: whether or not there exists an opposite word to pair with the noun in terms of gender and the nature of the lexical relationship between the alternants

Transparent:	gato (m)/gata (f)	'male cat, female cat'
Unmarked masculine:	jefe/jefa (m)/(f)	'male boss, female boss'
Lexical root gender:	hombre (m), mujer (f)	'man, woman'
Derivational:	cerezo/cereza (m)/(f)	'cherry tree, cherry'
Free variation (*alternating forms with minimal or no semantic change*):		
	canasta (f)/canasto(m)	'basket'
Unrelated:	plato (m)/plata (f)	'plate, silver'
Invisible gender (semantically alternating but marked only in article):		
	estudiante (f)/(m)	'male or female student'
Unique (only one gender per root):	carro (cf *carra)	'car'
Epicene (*one grammatical gendered semantically unspecified gender*):		
	víctima (f)	'male or female victim'

We finally considered both structural and referential context to assess whether the gender of a given DP was predictable. Predictability was categorical (y/n), and we annotated the source of predictability, either by syntactic (informative article) or by semantic/contextual means (known/previously mentioned). The purpose of this classification was to explore whether predictability of the gender of a DP predicts vowel underspecification.

Our final analysis combined the phonetic and the morphosyntactic results. The F1 and F2 values obtained for single word-final vowels in the narrative were Bark-transformed and combined with the morphosyntactic analysis.[8] In order to visualize the results, we created a new independent variable that was the result of subtracting the F1 (Bark) to the F2 (Bark). This allowed us to compare the degree of vowel centralization for each participant (the smaller the number, the more centralized the vowel). Then, we plotted F2 and F1 against the predictability of each noun and displayed the results by vowel organized by group. This visualization allowed us to explore the hypothesis of whether nouns that entered into predictable alternations showed a smaller degree of centralization than nouns whose gender is not predictable. To determine whether predictability played a role in vowel centralization, we ran linear mixed effects models with F2-F1 (in Bark) as the dependent variable, Vowel (/a e o/), Predictability (yes/no) and Group (Monolinguals, EB, LB) as independent variables, and Participant as a random factor. All statistics were calculated with R Studio Team (2020).

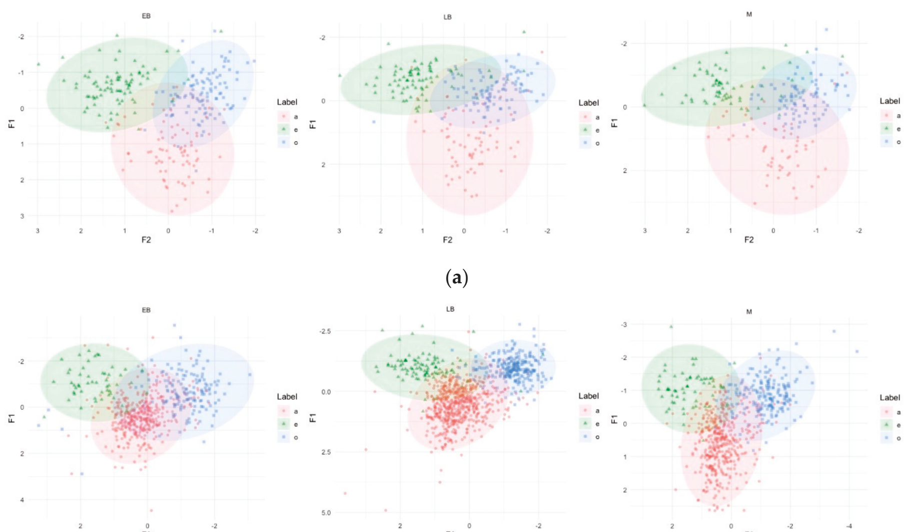
[8] The combined analysis only included data from the narrative because we are interested in comparing vowel quality and gender accuracy. The reading task, instead, is not clearly reflecting participants' grammatical knowledge.

3. Results

We begin this section by summarizing the results of the phonetic analysis, which includes the characterization of single vowels and vowel sequences. Section 3.2 presents the morphosyntactic analysis and the final section combines the phonetic results obtained for single vowels with morphological predictability to explore whether it may be possible that vowel centralization is being undetected when we report accuracy in gender agreement.

3.1. Phonetic Analysis

Figure 1a displays the F1–F2 normalized values obtained for single vowels in the reading task, whereas Figure 1b shows the results for the narrative. In both graphs, we observe some overlap, particularly between the vocalic spaces for /a/ and /o/ for all three groups. In the reading task, although bilinguals show slightly more overlap than monolinguals, the patterns are rather similar across groups. In Figure 1b, instead, both bilingual groups display a smaller vowel space for /a/, and EBs clearly show a greater degree of overlap between /a/ and /o/ realizations. This is important because these are the two vowels that are frequently used to encode gender.



(**a**)

(**b**)

Figure 1. Formant charts for /a e o/ in all groups (Early Bilinguals (EB), Late Bilinguals (LB) and Monolinguals (M)): (**a**) reading task; (**b**) narrative.

Table 2 summarizes two measurements of overlap; the Bhattacharyya's affinity scores and the percentage of overlap, calculated using convex hulls for the data obtained from the reading passage, whereas Table 3 displays the results of the narrative. As indicated in Section 2.3, both are complementary measurements to quantify overlap. Whereas the affinity score quantifies how similar the vocalic spaces are, the convex hulls quantify the degree of overlap. In both cases, the larger the number, the greater the overlap. As mentioned, the goal of these measurements is to quantify overlap, not to test statistical significance among groups.

These results show, first, a task effect and, second, a vowel-pair effect. The degree of overlap across pairs of vowels (Table 2) is lower in the reading task than in the narrative, and this is particularly evident in the values calculated using convex hulls. More careful articulation and, thus, less overlap is generally expected in a reading task. The degree of overlap, instead, increases in the narrative at different rates across groups (Table 3). LBs and particularly EBs double and sometimes triple (e.g., [a]–[o]) the degree of overlap in vocalic spaces when speaking spontaneously (see overlap %), which suggests that the distinction between some pairs of vowels may be weakening. This is clearly the case with the [a]–[o] pair, and, to a lesser extent, with the [a–e] pair.[9] In turn, this tells us that the most frequent vowel pairs that mark gender are being produced with the same quality a third of the time by EBs.

Table 2. Affinity scores and proportion overlap (convex hulls) for /a e o/ in the reading task. Results displayed by group (Early Bilinguals (EB), Late Bilinguals (LB) and Monolinguals (M)).

Group	Affinity (0–1)			(Convex Hulls) Overlap (%)		
	a–e	a–o	e–o	a–e	a–o	e–o
EB	0.23	0.49	0.21	13.8	10.2	11.8
LB	0.35	0.48	0.28	15.4	11.9	11.5
M	0.31	0.40	0.36	15.8	10.5	9.8

Table 3. Affinity scores and proportion overlap (convex hulls) for /a e o/ in the narrative. Results displayed by group (Early Bilinguals (EB), Late Bilinguals (LB) and Monolinguals (M)).

Group	Affinity (0–1)			(Convex Hulls) Overlap (%)		
	a–e	a–o	e–o	a–e	a–o	e–o
EB	0.36	0.52	0.17	26	32.1	21.3
LB	0.35	0.24	0.17	27.4	25.1	15
M	0.36	0.41	0.16	17.5	18.9	17.8

We turn to the realization of sequences, which were analyzed independently from single vowels because the formant quality and trajectory largely depends on how they are realized. If realized as hiatuses, formant values should be similar to those obtained for single vowels, whereas values in diphthongs should differ from the values of the corresponding single vowels (e.g., Borzone de Manrique 1976; Aguilar 1999). Thus, before conducting the acoustic analysis, it is important to determine the proportion of sequences produced together from those realized in different syllables (Figure 2). A preference for hiatuses will be interpreted as indicative of cross-linguistic influence, since across-word vowels in Spanish tend to be pronounced together (e.g., Aguilar 2010), which is not the case in English (Davidson and Erker 2014). The fact that vowels are syllabified differently across words in Spanish has also important implications for L1 acquisition because, as mentioned, the vowel quality in diphthongs differed from the vowel quality in singleton vowels. Thus, a higher probability of diphthongs is also a higher probability of noise in the signal for bilinguals and, thus, for difficulty in determining the quality of the vowel that marks gender. Results in Figure 2a show that EBs, in the reading task, had a larger proportion of hiatuses than the other groups, whereas all groups behaved similarly in the narrative (Figure 2b).

To determine if there were differences in the realization of sequences labeled as diphthongs, we compared the formant trajectories across groups.[10] First, we analyzed the sequences produced by female speakers in the reading task. Male speakers' productions were not analyzed given the small number of tokens obtained for this task (see Methods) and the even smaller number of tokens in which vowels were not separated by a pause or glottal stops. In terms of F1 trajectories, no statistical differences were observed across the three groups (Appendix A, Figure A1). In terms of F2 trajectories, the SSANOVAs revealed significant differences for all sequences, where the EB group realized (i) /ae/

[9] Results regarding the vowel [e] in the narrative should be interpreted with caution, because this vowel represents only 10% of the total tokens.
[10] Note that only sequences that were realized as diphthongs in all groups were analyzed.

with significantly lower values than M and LB (Figure 3a), (ii) /oe/ with significantly lower values than LB (Figure 3b), and (iii) /ea/ with significantly lower values than M and LB (Figure 3c). The F2 trajectories realized by the M and LB groups did not differ significantly.

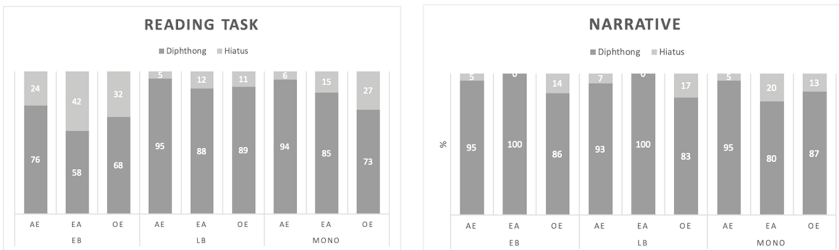
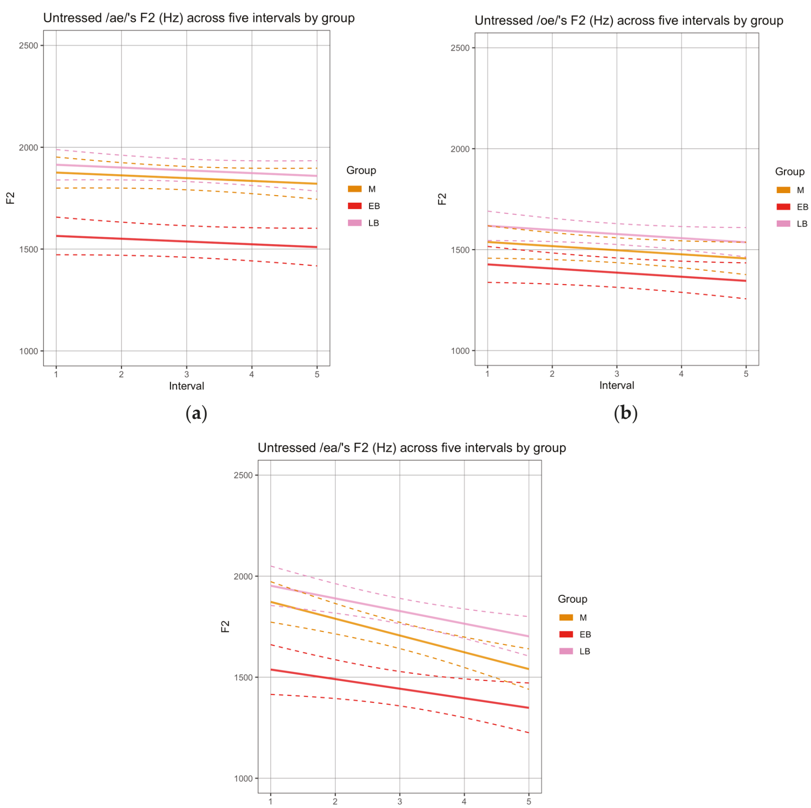


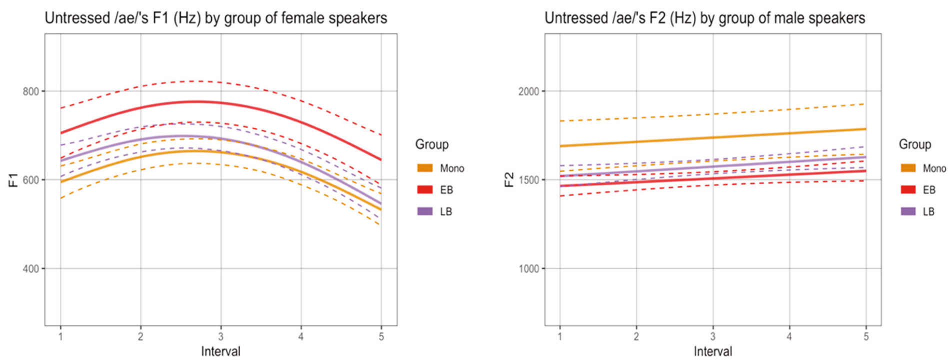
Figure 2. Percentage of diphthongs and hiatuses by type of vowel sequence and group (horizontal axis): (**a**) Reading task; (**b**) narrative.



(**c**)

Figure 3. Smoothing Splines ANOVAs (SSANOVAs) for non-normalized (female) F2 trajectories across five intervals produced by each group of speakers (M, EB, LB) in the reading task. (**a**) Sequence /ae/; (**b**) sequence /oe/; (**c**) sequence /ea/.

Second, and in order to be able to compare results across tasks, we analyzed the sequences produced by female and male speakers in the narrative separately. SSANOVAs by gender using non-normalized values revealed two statistical differences, both within the sequence /ae/: (i) the female EB group realized F1 trajectories with significantly higher values than the female M group (Figure 4a); and (ii) the male EB group realized F2 trajectories with significantly lower values than the male M group (Figure 4b).[11]



(**a**) (**b**)

Figure 4. SSANOVAs for non-normalized formant trajectories across five intervals within the sequence /ae/ produced by each group of speakers (M, EB, LB) in the narrative task: (**a**) F1 trajectories: female speakers; (**b**) F2 trajectories (male speakers).

Taken together, these results suggest that the EB group behaves differently from the M and LB groups, often centralizing F2 values in unstressed vowel sequences.

3.2. Morphosyntactic Analysis

The grammatical analysis showed high accuracy counts. From a total of 2445 DP tokens extracted for analysis, we identified six clear gender errors (see (7)–(12) below), as well as another handful of errors with grammatical number, mostly singular in lieu of plural, which will not be discussed here. Erroneous marking is indicated in bold.

(7) la abuelita y la caperucita roj**o** . . . (Speaker UT052, EB)
The-FEM grandmother-FEM and the-FEM little-riding-hood- FEM red-MASC

(8) Y la niña la caperucita roj**o** . . . (Speaker UT052, EB)
And the-FEM girl-FEM the-FEM little-riding-hood-FEM red-MASC

(9) con **ello** y a a le hablaba y . . . (Speaker UT054, EB)
With **him** and ah ah her-DAT spoke and [intended reference to "with **her**"]

(10) bueno su abuel**o** la había . . . (Speaker UT107, M)
Well her grandparent-MASC her-DAT had

(11) todas l**as** alimentos . . . (Speaker UT086, LB)
All-FEM the-FEM food-MASC

(12) tant**o** zozobra . . . (Speaker UT002, LB)
So-much-MASC unstability-FEM

We observed two additional, more ambiguous instances where a demonstrative could function as a filler or a stranded phrase or a demonstrative. In (13), below, we note that Speaker UT030 was a

[11] Non-significant results for female and male speakers are included in Figures A2 and A3, respectively.

frequent user of the *este* filler. Nonetheless, there was no prosodic break separating the noun from the preceding demonstrative, so an analysis of these tokens as instances of determiner-noun disagreement cannot be ruled out.

(13) est**e** caperucit**a** está entrando . . . (Speaker UT030, M)
This-MASC little-riding-hood-FEM is entering

(14) y es**o** cosas como eso . . . (Speaker UT054, EB)
And this-MASC things-FEM like this-MASC

This high accuracy rate is appropriate, given the high proficiency status of these speakers, many of which produced elaborate, lexically rich and syntactically complex narratives that were, for the most, seemingly error free and without English intrusion. Nonetheless, the formal and contextual properties of gender expression in the DP deserve further examination. We ask two questions: How robustly do these DPs manifest gender agreement explicitly? And given our phonetic results, how accurate will a coder or transcriber be at detecting that a given gender vowel has been centralized. If a gender form is highly predictable (given lexical retrieval of a contextually or syntactically predictable entry), a coder might be likely to perceive a centralized vowel as the target. Given what is known about speech perception, we can expect that any skilled listener will predict a centralized vowel (caperucit@) to be an /a/ and ignore the presence of the schwa. Even if we think we are paying attention, the evidence suggests otherwise. Listeners are generally known to perceive elements that are not present in the speech chain in order to repair phonotactically illicit sequences (e.g., Calabrese 2012; Durvasula and Kahng 2015a; Durvasula and Kahng 2015b; Hawkins 2010; Repp 1992); so, we might expect them to easily attribute features to underspecified central vowels.

To further explore these patterns of accuracy, we consider the distribution of lexical and morphological types of noun, determiner, and adjective forms. Nouns were fairly evenly divided between those that entered into a gender alternation (52%, 1284 tokens) and those that did not; and, for this particular story, feminine nouns were almost twice as frequent as masculine nouns (1575 vs. 885 tokens). For most nouns, gender marking is an arbitrary classification with no semantic import. The main exceptions are nouns referring to human entities, some but not all animal classes (cf. *gallina* vs. *avispa*), and a handful of narrow subclasses of systematic alternations (*cerezo/cereza*). Nouns with human referents (*abuela/madre/caperucita/niña/cazador*) made up 42% of the data; animals (mostly *lobo*, 'wolf', with an occasional reference to cats in Caperucita's house) added another 12% to the count of semantically transparent gender alternations.

From a lexical perspective, over 80% of the noun tokens (2028 tokens) in the narratives had explicit gender marking (i.e., transparent *-o/-a* word-markings). Other nouns had either opaque word-final morphology (ending in *-e* or consonant) or ended in a transparent suffix (*-ción*, etc.). A potential loss of the gender system is revealed either as (i) a switch in the word marker vowel (i.e., saying *caperucit**o*** or *abuel**o** in lieu of abuel**a***), or (ii) a mismatch between the noun and the determiner (*l**a**s aliment**o**s*). A determiner-noun mismatch can potentially indicate loss of concord or agreement; most likely, it may just represent a lexical knowledge gap: the speaker may have misclassified the gender class to which a noun belongs. A study of gender agreement in French-English bilingual children by Nicoladis and Marchak (2011) supports this claim. Their data showed less attrition in concord (i.e., agreement between determiner-adjective) than in gender assignment (i.e., what determiner was associated to a noun). Bilingual children were significantly less accurate with gender assignment beyond their differences in lexical scores. However, when these authors strictly considered concord, there were no statistical differences between monolingual and bilingual children.

Determiners as a category were only partially informative. Bare nouns were common (26%, 629 tokens), and some determiners such as possessives (*mi casa*) and quantifiers (*dos casas*) were uninformative for gender (14%, 347 tokens), so that about 40% of the analyzed tokens had no gender identification in the determiner. The determiner forms that inflect for gender (*definites, indefinites, demonstratives*) can be divided into those where gender identifiability depends only on perceptibility of the gender vowel, as for demonstratives (*ese niño/esa niña*) or the plural cases of definites and indefinites. Only for the singular definite and indefinite determiners (*un caso/una casa*) is there additional phonetic information beyond vowel quality, provided the following word starts with a consonant (i.e., *una gata* but not *una_abuela*). Table 4 shows the frequencies of all determiner types classified by number in the noun. The frequencies of the more informative singular definite and indefinite forms are indicated in bold.

Table 4. Distribution of DPs by determiner and number (all tokens).

Determiner	Plural	Singular	Total
def	68	1011	1079
dem	35	52	87
indef	20	279	299
null	94	535	629
poss	13	292	305
quant	15	27	42
wh	10	7	17
All DPs	255	2203	2458

If we consider only the determiners marked for gender (definite/indefinite/demonstratives) for nouns ending with word markers -a, -o, we are left with 1156 or 47% of the nouns. On the assumption that a speaker or listener would ignore vowel quality for non-alternating nouns ([bok@] for *boca*) and only control production/perception of contrasting vowels, where it matters ([niñ@] would have to be resolved into *niño* or *niña*; similarly for [kas@] *caso/casa*), we could restrict our attention to only alternating nouns. As shown in Table 5, only 791 tokens or 32% of all data has nouns that will mark gender and are accompanied by a determiner with visible gender.

Table 5. Frequencies of determiner types associated with gendered nouns (reported for all gendered nouns and separately for only those that entered into an alternation).

Determiner	All Gendered Nouns		Only Alternating Nouns	
	Plural	Singular	Plural	Singular
Definite	50	838	10	651
Indefinite	15	184	6	111
Demonstrative	25	44		13

So far, we have only talked about form. The scenario is even worse if we take into account the specific context of this narrative and deem as highly detectable nouns that are not contextually predictable. As pointed out above, if someone murmurs *Caperucito roj@* we are likely to hear *Caperucita roja* for the simple reason that *Caperucita* is one and only. In our story of choice, there are five main characters; all of them highly contextually predictable. We know we are talking about an *abuela*, not an *abuelo*, and that there is no *loba* in the story. We first note that there is a strong association between semantic type of noun and contextual status, as shown in Table 6. This is in the direction one would expect, with the five main characters (animate, the wolf and humans, *Caperucita*, grandmother, mother and hunter) which make up the bulk of the contextually-given reference. Table 6 cross-tabulates all nouns; columns classify all nouns by whether there was a preceding article or not, and rows show how those nouns are distributed in relation to contextual information, separated by whether it consisted of previous mention, first initial mention of known entities, or neither type of identification. We restricted previous mention to subsequent repetition using the same lexeme; that is, saying *Caperucita* twice was counted as previous mention but not *la niña*).

Table 6. Cross-tabulation of semantic type of nouns against contextual status of the DP.

Contextual	Concrete	Abstract	Event	Anim	Hum
Previous mention	19			243	755
Known	2			29	141
Neither	931	163	3	29	129
Grand Total	952	163	3	301	1025

This leaves us with 24 gendered, alternating nouns that were not gender-predictable, neither on the basis of syntax (preceding article) nor context.

3.3. Combined Results

The last step in this exploration is to combine what we found about the phonetic realization of these vowels and the morphological analysis. Thus, we will discuss here a way of combining the phonetic results obtained from the narrative elicitation task (the results of the reading task are left aside because read speech is not reflective of the underlying grammatical system) with the analysis of predictability presented in Table 7. Given the characteristics of our dataset (unbalanced number of vowels per category, distribution of predictable and unpredictable nouns both in the story and across speakers), this is very much a tentative proposal rather than a strong claim.

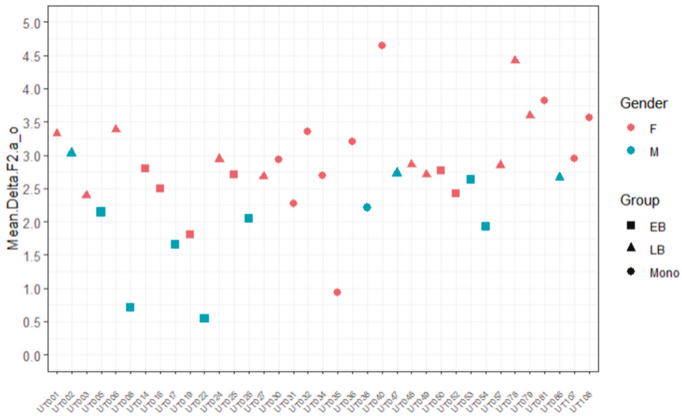
Table 7. Cross-tabulation of nouns (reported separately for all nouns and for gendered nouns in alternation) against contextual status of the DP.

Context	All Nouns		Only Gendered Nouns in Alternation	
	Preceding Article	No Syntactic Cues	Preceding Article	No Syntactic Cues
Previous mention	553	465	459	7
Known	80	92	40	1
Neither	669	586	256	24

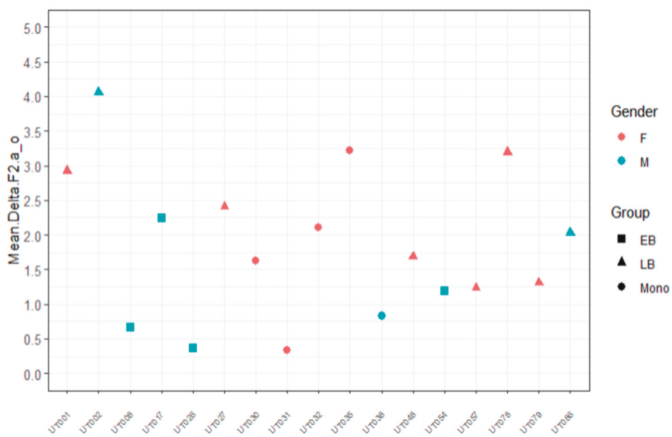
As discussed in Section 2.3, to combine our results, we Bark-transformed the F1 and F2 values and calculated the F2–F1 difference to obtain a single result for each token. As a reminder, a small difference in Bark is interpreted as a sign of centralization. Our first analysis, displayed in Figures 5 and 6, presents the summary statistics for /a o/.[12] For each vowel, we calculated the mean difference for predictable and non-predictable nouns. Then, to each data point in our Excel file, we subtracted the mean value obtained for the opposite vowel with the same degree of predictability. For example, for a given token of the vowel [a] in *Caperucita*, we subtracted the mean obtained for all the [o] vowels in words like *lobo*.[13] If participants are making no difference between these unstressed vowels, the numbers should be closer to 0.

12 We focus on these two vowels, given that we had a larger number of tokens than for /e/ and that 80% of the nouns that were marked for gender in the story had these vowels.

13 The graphs also reflect the fact that there were many more predictable nouns than non-predictable nouns in this narrative, hence the difference in the number of data points in Figures 5a and 6a when compared to Figures 5b and 6b.



(**a**)

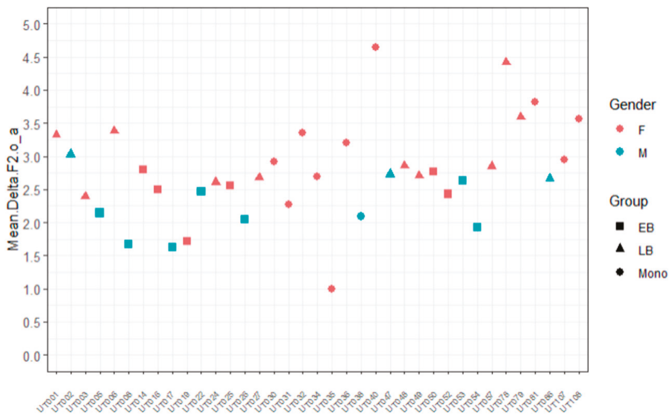


(**b**)

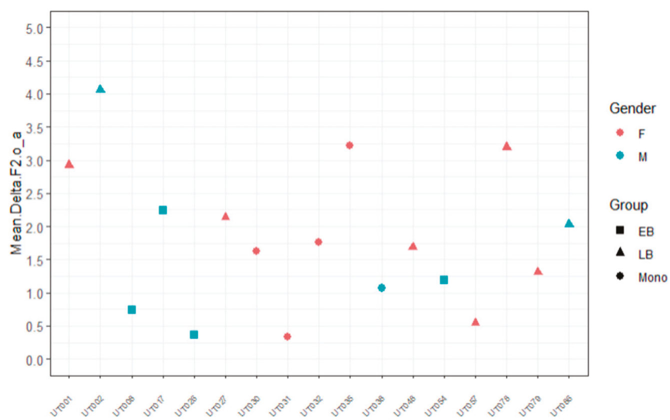
Figure 5. F2-F1 in Bark for the vowel /a/: (**a**) Values obtained for /a/ minus values obtained for /o/ in predictable contexts; (**b**) values obtained for /a/ minus values obtained for /o/ in non-predictable contexts.

Figure 5 shows that values obtained for EBs are closer to 0 (mean in predictable contexts = 1.95; mean in unpredictable contexts = 0.94) than for the other two groups (LB_predictable = 3.09; LB_unpredictable = 2.53; mono_predictable = 3.02; mono_unpredictable = 1.99), which means that EBs make less of a distinction between the two vowels. Values obtained for /o/ (Figure 6) display the same tendency, albeit mean values per group are slightly higher than those obtained for /a/.[14] The analysis of both vowels then suggests that the patterns are the opposite to those that we had hypothesized; namely, the difference between the two contrasting vowels was larger in predictable than in non-predictable contexts.

[14] EB_predictable = 2.15; EB_non-predictable = 1.33; LB_predictable = 3.08; LB_non-predictable = 2.43; mono_predictable = 2.91; mono_non-predictable = 1.80.



(**a**)



(**b**)

Figure 6. F2-F1 in Bark for the vowel /o/: (**a**) Values obtained for /o/ minus values obtained for /a/ in predictable contexts; (**b**) values obtained for /o/ minus values obtained for /a/ in non-predictable contexts.

The second type of analysis conducted on this combined dataset was a series of mixed effects models with the F2–F1 (in Bark) difference for each vowel token as the dependent variable and Vowel (/a e o/), Language group (M, EB, LB) and Predictability (yes, no) as the independent variables. Reference levels are /a/ for Vowels, M for Group, and no for predictability. Table 8 reports estimates, standard error, and significance of a linear mixed effects models with Participant as a random effect and Vowel, Language Group, and Predictability of the noun gender as fixed effects (Winter 2020).

The combined analysis of morphosyntactic and phonetic results presented in this section reveals that all groups centralize vowels more in gender-predictable nouns. Further analysis of predictability by formant distance (Figures 5 and 6) suggests that EBs showed a higher degree of centralization in /a o/ than the other groups.

Table 8. Results of a linear mixed effects model for the F2–F1 (Bark) with Participant as a random effect and Vowel, Language Group, and Predictability of the noun gender as fixed effects. Reference values: Vowel = [a]; Language Group = Monolinguals; Predictability = no.

Fixed Effects	Estimate	*SE*	*t*	$p(>\|t\|)$
Intercept	5.05	0.18	27.52	<0.001
Vowel(e)	−3.03	0.12	24.16	<0.001
Vowel(o)	−0.97	0.07	−13.38	<0.001
EBs	−0.10	0.22	0.047	0.63
LBs	−0.11	0.22	−0.50	0.61
Predictability(yes)	0.17	0.09	1.94	0.05

4. Discussion

4.1. Hypothesis Evaluation

Our first research question and hypothesis targeted the phonetic realization of word-final unstressed singleton vowels and vowel + vowel sequences. We hypothesized that EBs would show a higher rate of overlap in the vocalic spaces of the three target vowels than the other two groups. Results showed this to be the case, particularly in the narrative. Figure 1b and Table 3 point to the fact that the vocalic spaces for the vowel pair that encodes gender more often (i.e., [a]–[o]) overlap in more than 30% of the tokens. This means that 1/3 of these vowels are realized with the same quality. We also hypothesized that monolinguals and LBs would tend to fuse vowels across words while EBs would tend to separate them. Figure 2 showed this to be the case only in the reading task. We further explored the quality (formant trajectories) of the sequences that were pronounced without a pause or a glottal stop to determine potential group differences that could be attributed to influence from English. Given the smaller number of tokens obtained for sequences when compared to single vowels and that most of those tokens were produced by female speakers, we focused on the latter in our statistics for the reading task. We showed that EBs indeed displayed signs of centralization in the realization of such sequences, since /ae/ and /ea/ were realized with similar trajectories and with lower F2 values than those obtained for the other groups. For the narrative, we analyzed both normalized and non-normalized values produced by female and male speakers. Although no statistical differences were observed in the former analysis, the latter revealed signs of centralization in EBs' realization of the sequence /ae/, which was produced with lower F2 values by male speakers and higher F1 values by female speakers than by the monolingual counterparts.

Our second question concerned the perceived errors in gender agreement and concord. Based on previous literature, we expected to find more errors in the EB group than in the other groups. However, very few errors were found in total (only six) and half of them were produced by EBs, which does not seem to support our hypothesis.

Our third research question was explored with data obtained from the narrative and addressed the issue of potential underreporting of gender errors. We hypothesized that vowel centralization would be labelled as accurate gender. Based on the results obtained for the phonetic characterization of vowels (we found 32% of overlap in the vocalic spaces for [a o] in EBs) and the fact that only three errors (examples 7–9) were perceived in this group, we conclude that the hypothesis was confirmed. As we will discuss in the next section, we would have expected at least a higher number of cases labeled as questionable but we ourselves did not do that.

Our final question concerned the results obtained in the combined phonetic and morphosyntactic analysis. Here, we wanted to determine whether EBs would be more likely to mark gender (i.e., both accurate marking and less vowel centralization) in DPs that were not predictable, neither syntactically nor contextually. Our results rejected our hypothesis that this would be the case. Indeed, we found that all groups (and especially EBs) had less centralization in predictable than in non-predictable contexts.

In the next section, we discuss why this may be the case and we contextualize our results in terms of past and future research.

4.2. General Discussion

The results of our phonetic analysis show that there are differences in the realizations across groups and these differences are larger in the narrative than in the reading task. In particular, we have shown that EBs tend to centralize /o/ and have, as a consequence, a high degree of overlap in the vocalic space for [a]–[o]. This is consistent with previous production studies (Gildersleeve-Neumann et al. 2009; Menke 2010; Rogers 2012; Ronquest and Rao 2018) that reported differences in the realization of single vowels between heritage speakers and Spanish monolinguals. The patterns reported here are also consistent with previous perception studies showing that L1 English–L2 Spanish speakers tend to confuse /o/ with /a/ (Morrison 2003; Morrison 2006) and that Spanish heritage speakers tend to confuse vowels but only in unstressed positions (Mazzaro et al. 2016). We have also seen that across-group differences are not restricted to single vowels; there were also differences in the realization of vowels across words. EBs had a higher proportion of hiatuses than the other groups in the reading task and had differences in the F2 trajectories in all the sequences analyzed (i.e., /ae oe ea/). Differences were also found in the narrative for the F1 (female speakers) and F2 (male speakers) trajectories in the /ae/ sequence. Overall, our results showed signs of centralization of the vocalic space, signaled by changes in the F2 rather than in the F1. There are two important points to keep in mind. First, we have found differences in EBs, when compared to the other two groups, in a community with a high level of bilingualism and where Spanish is omni-present. All EBs had or were receiving some education in Spanish and used Spanish daily. They also showed DELE scores very close to those obtained by LBs. Second, we are reporting here results of an unbalanced dataset. Participants are not equally distributed between gender groups (which was overcome in part with normalization) and did not contribute to the sample the same number of tokens (narrative). Although everybody produced the same number of vowels in the reading task, our main analysis refers to the narrative, to which each participant contributed a different number of tokens distributed across different lexical items.

The analysis of gender agreement and concord revealed very few errors, and half of these errors were generated by the early bilinguals. Beyond these few errors, we extracted three observations about the distribution of morphological types. These observations inform our understanding of the challenge of assessing the distribution of gender realizations. First, while most nouns (4 out of 5) are explicitly marked, only a subset of all nouns are accompanied by transparent determiners (3 out of 5). Second, for all transparent nouns and for most transparent determiners, the single cue to gender realization is the vowel: a centralized vowel might actually be ambiguous for accuracy, but vowel centralization is likely to remain undetected, particularly in predictable, non-contrastive contexts. Last, the vast majority of the NPs we analyzed were highly predictable in context and almost never contrastive.

If we now turn to our combined results, we showed in Table 3 that, in 32% of the /a o/ tokens produced by EBs, the vocalic spaces completely overlapped. That means that there should be 221 tokens (out of the 737 tokens of these two vowels) in which the transcriber should have expressed some doubts about the quality of the vowel. If we turn to Section 3.2, we see that we only labeled six cases as gender mismatch. Thus, one can say that, in coding our data, we might have over-reported gender accuracy in possibly as many as 215 tokens, and this only in reference to one of the vowel pairs. If we turn to the combined analysis, both Figures 5 and 6 showed that EBs tend to make smaller distinctions between /a/ and /o/, both in predictable and in non-predictable contexts. To sum up, despite the limitations inherent to this type of data (narrative elicitation) and to our specific dataset (*La Caperucita*), we believe that we have established preliminary grounds for what we have dubbed the *modular interaction hypothesis*, which proposes that changes in one domain (phonetics) can have consequences for the acquisition of other domains (morphosyntax). We have made a case for phonetics as both a methodological obstacle for assessment of bilingual gender grammars as well as a potential incipient factor in contact-induced grammar restructuring in bilinguals.

The results of the combined analysis also showed a main effect of predictability of the noun gender; namely, nouns that were predictable had a significantly higher F2–F1 difference than unpredictable nouns. This means that the distance between the formants, which is a proxy of vowel centralization, was higher in predictable nouns, contrary to our last hypothesis. However, if we return to the data in Table 6, we note that Predictability and Animacy were heavily overlapping categories; i.e., all our predictable nouns were animate, while few inanimate nouns were predictable. This offers an alternative interpretation of the data as signaling that participants are enhancing the marking of gender with animate nouns. Current data does not allow us to evaluate these two possible interpretations of the data.

We do not seek to emphasize the limitations of narrative or spontaneous data, or of previous studies, nor to argue that there are or might be more gender errors than those reported in the literature and/or in the present results. Our goal was to assess how far bilingual speakers can go in reducing the quality of unstressed final vowel contrast before other speakers (including transcribers) detect that something is different or missing in the signal. On the route to this important point, we have shown that, although all DPs are analyzed as possessing abstract gender features, only a portion of them visibly show it. Equally important, only a very small subset of these realized gender forms is not predictable from the syntax or the semantics. As such, these results suggest that future studies of bilingual gender marking should address phonetic analyses explicitly.

The implications are more important than a methodological point. We have examined narrative use of gender and discovered that, in general, gender in many nouns is not overtly marked, but that gender is often a category that can be easily predicted from context. We have also shown evidence, as we predicted, that bilingual heritage adults are likely to produce agreement vowels that are less distinctly specified than those that will be available in the environment of a child in a monolingual community. In a bilingual context with an ongoing contact-based phonetic shift, the signal to gender marking becomes increasingly opaque for the child learner. This contact-induced variation includes robust patterns of vowel centralization and a high degree of overlap in the vocalic space. Furthermore, we should not leave aside the variation that results from the co-articulation between word-final vowels and words beginning with vowels in monolingual speech. Our bilinguals are separating vowel sequences in their production; this means that fused sequences in the speech of monolinguals are likely to remain opaque for them in perception.

Given this variability in the input, we speculate that, at some point, gender in contact varieties of Spanish could become French-like. That is, a scenario where gender is still marked in the syntax but is more visible in the determiner than in the noun morphology. The possibility remains that, for some children, the disruption in the acoustic signal may lead to the development of systems where there has been a reorganization of the underlying featural system, or even a more extreme scenario where gender is no longer part of the agreement system and remains as a vestigial lexical remnant. It is worth recalling the results in Cuza and Pérez-Tattam (2016) data, where, as in other studies, there are more errors with feminine targets than with masculine targets. The individual analysis of their data shows that about a quarter of the children produced no single instance of correct feminine agreement. Those children give the appearance of having contracted the system to the masculine default. Again, more data, particularly data with individual analyses, is needed to explore the range of morphological reorganization processes that might characterize individuals in the North American bilingual context, and to study the potential link between morphosyntax and changes to vowel perception and representation.

Our current study allows us to re-contextualize some of the pre-existing literature. As sociolinguists hold, narratives are a useful tool to elicit semi-spontaneous speech. The narrative analyzed in this paper has special characteristics; namely, it is a traditional folk tale widely known in the Western world. Thus, many lexical items are highly predictable, but not all. In Section 3.2, we showed that there is actually a very small proportion of nouns whose gender is not predictable. This suggests that, in most cases in real life, a bilingual can consistently communicate inserting schwas in word-final position without them being detected. This, in turn, could be part of the observed asymmetries in results reported in studies such as Montrul et al. (2014), in which participants are accurate in production tasks but inaccurate in grammaticality judgement tasks. The analysis of the narrative, in combination with the study of the phonetic realization of vowels, shows that there are small but systematic differences in the speech of EBs when compared to the other groups. These small differences probably go undetected in production when these speakers interact with other bilingual and monolingual speakers but are substantial enough to limit their perceptual skills and possibly, for some bilingual speakers, prevent them from attaining the underlying representational system.

Our use of semi-spontaneous, narrative data introduced asymmetries and gaps in the data, which precluded certain analyses. At the same time, it gave us a better view of natural speech patterns than the reading data and allowed us to contextualize morphosyntactic analyses and moderate the conclusions from those analyses in light of the phonetic findings.

5. Conclusions

We have shown that early bilinguals differ in the phonetic realization of vowels when compared to monolinguals, as has been reported in previous studies. We argued that those differences have implications for the way in which we should analyze agreement errors (or the lack of them) in Spanish-English bilinguals. Even in our population of early bilinguals, who use Spanish daily, we have seen, on the one hand, low rates of reported errors in agreement, as identified by human coders, and on the other, frequent patterns of vowel centralization and a high degree of vowel overlap that cast doubt on informal analyses of gender marking. We argue that such phonetic reality provides the basis for established patterns of incipient restructuring in the morphological system of Spanish heritage bilinguals (as suggested by Scontras et al. 2018) as well as other, not yet fully documented possible impacts to the system.

Our final take home message is simple. Misperception happens for single segments (e.g., Calabrese 2012; Durvasula and Kahng 2015a; Durvasula and Kahng 2015b; Hawkins 2010; Repp 1992; see also Ohala 1989, 1993) as well as for whole utterances. Speaking about mondegreens, i.e., misheard song lyrics, Nevins (2014) explains: "listeners might impose what they wish to hear in the songs, or perhaps, what they expect to hear, in perception". Spanish-speakers predict and expect grammatical gender, so they hear gender vowels when, in fact, the bilingual speaker has spoken a schwa. Thus, it is important to cross-analyze data across linguistic domains and to remember that we linguists, as all human beings, believe we hear what we already know.

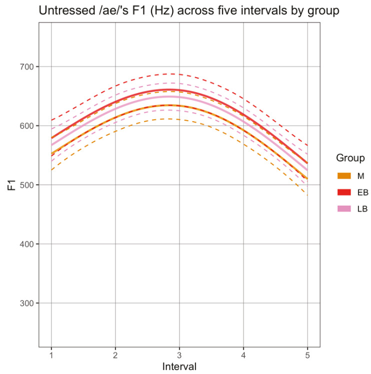
Author Contributions: Conceptualization: L.C. and A.T.P.-L.; methodology and analyses: all contributors; data collection: N.M.; writing: all contributors; funding acquisition: L.C., N.M. and A.T.P.-L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by a Social Science and Humanities (SSHRC) Insight grant from the Government of Canada, grant number 435-2020-0110.

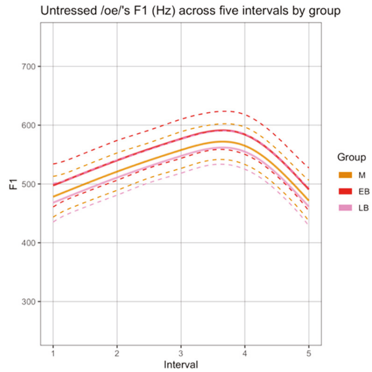
Acknowledgments: We thank our participants for their time and M. Barreto and M. Lazzari for their assistance with the statistical analysis.

Conflicts of Interest: The authors declare no conflict of interest.

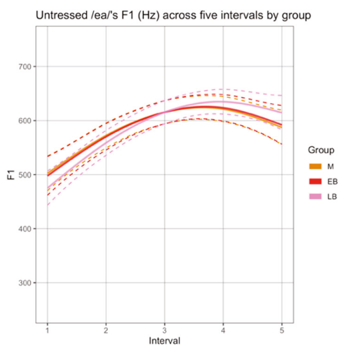
Appendix A



(**a**)

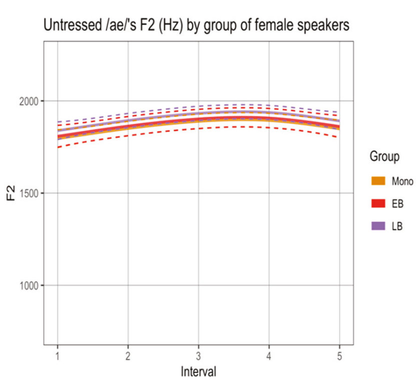


(**b**)

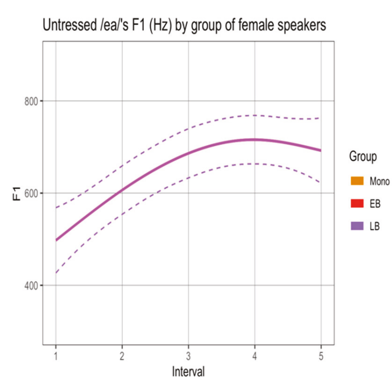


(**c**)

Figure A1. SSANOVAs for non-normalized (female) F1 trajectories across five intervals within the sequences (**a**) /ae/; (**b**) /oe/; and (**c**) /ea/ produced by each group of speakers (M, EB, LB) in the reading task.

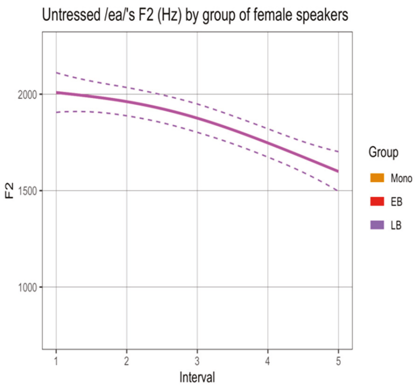


(**a**)

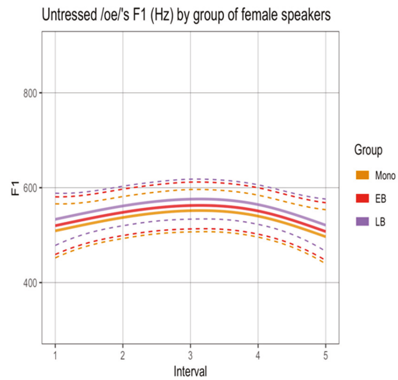


(**b**)

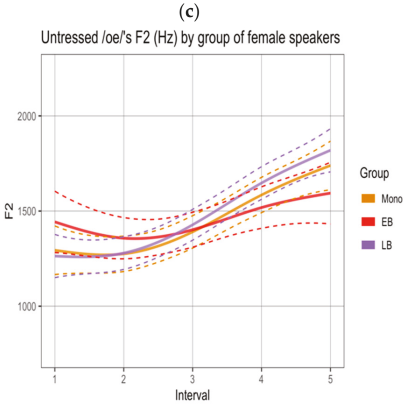
Figure A2. *Cont.*



(**c**)

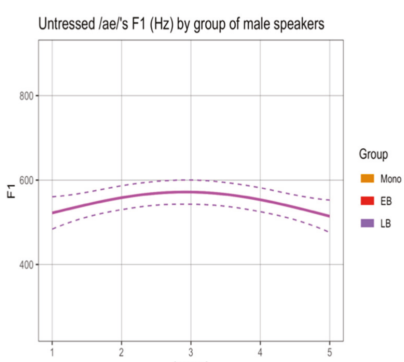


(**d**)

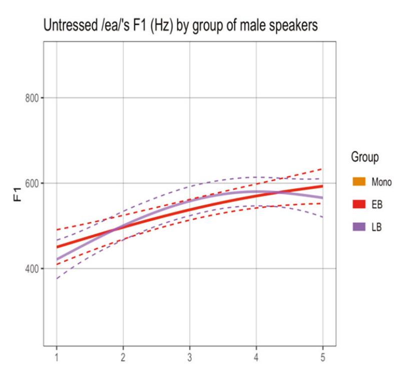


(**e**)

Figure A2. SSANOVAs for female fromant trajectories across five intervals within the sequences (**a**) /ae/ (F2); (**b**) /ea/ (F1); (**c**) /ea/ (F2); (**d**) /oe/ (F1); and (**e**) /oe/ (F2) produced by each group of speakers (M, EB, LB) in the narrative task.

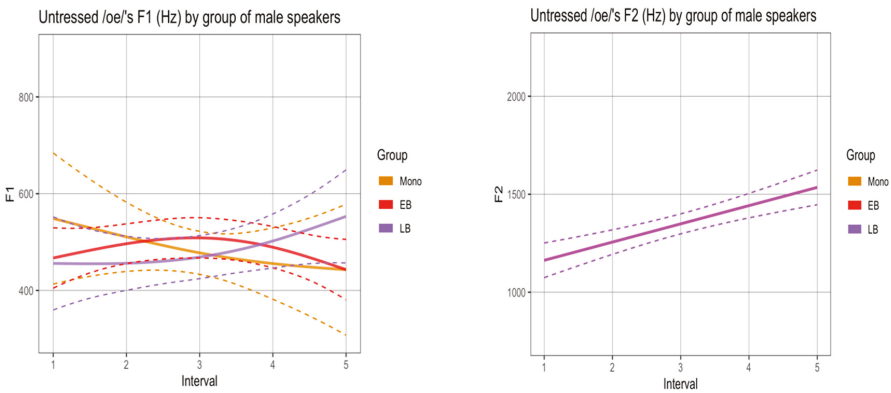


(**a**)



(**b**)

Figure A3. *Cont.*



(**c**) (**d**)

Figure A3. SSANOVAs for male formant trajectories across five intervals within the sequences (**a**) /ae/ (F1); (**b**) /ea/ (F1); (**c**) /oe/ (F1); and (**d**) /oe/ (F2) produced by each group of speakers (M, EB, LB) in the narrative task.

References

Adank, Patti M., Roel Smits, and Roeland van Hout. 2004. A comparison of vowel normalization procedures for language variation research. *Journal of the Acoustical Society of America* 116: 3099–107. [CrossRef]

Aguilar, Lourdes. 1999. Differences. *Speech Communication* 28: 57–74. [CrossRef]

Aguilar, Lourdes. 2010. *Vocales en Grupo*. Madrid: Arco/Libros.

Alba, Matthew C. 2006. Accounting for variability in the production of Spanish vowel sequences. In *Selected Proceedings of the 9th Hispanic Linguistics Symposium*. Edited by Nuria Sagarra and Almeida. J. Toribio. Somerville: Cascadilla Press, pp. 273–85.

Barreña, Andoni. 1997. Desarrollo diferenciado de sistemas gramaticales en un niño vasco-español bilingüe. In *Contemporary perspectives on the acquisition of Spanish*. Edited by Ana T. Pérez-Leroux and W. Glass. Somerville: Cascadilla Press, vol. 1, pp. 55–74.

Bedore, Lisa M., and Laurence B. Leonard. 2001. Grammatical morphology deficits in Spanish-speaking children with specific language impairment. *Journal of Speech, Language, and Hearing Research* 44: 905–24. [CrossRef]

Bernhardt, Barbara H., and Joseph P. Stemberger. 1998. *Handbook of Phonological Development*. San Diego: Academic Press.

Boersma, Paul, and David Weenink. 2001. Praat: Doing Phonetics by Computer [Computer Program]. Available online: http://www.praat.org/ (accessed on 15 June 2020).

Borzone de Manrique, Ana M. 1976. Acoustic study of /i, u/ in the Spanish diphthongs. *Language and Speech* 19: 121–28. [CrossRef]

Calabrese, Andrea. 2012. Auditory representations and phonological illusions: A linguist's perspective on the neuropsychological bases of speech perception. *Journal of Neurolinguistics* 25: 355–81. [CrossRef]

Calenge, Clément. 2006. The package adehabitat for the R software: A tool for the analysis of space and habitat use by animals. *Ecological Modelling* 197: 516–19. [CrossRef]

Castilla, Anny P., and Ana T. Pérez-Leroux. 2010. Omissions and substitutions in Spanish object clitics: Developmental optionality as a property of the representational system. *Language Acquisition* 17: 2–25. [CrossRef]

Colantoni, Laura, and José. I. Hualde. 2016. Constraints on front-mid vowel gliding in Spanish. In *The syllable and Stress*. Edited by R. Nuñez Cedeño. The Hague: Mouton de Gruyter, pp. 1–28.

Colantoni, L., and A. Limanni. 2010. Where are hiatuses left? A comparative study of vocalic sequences in Argentine Spanish. In *Selected Proceedings of the 38th Linguistic Symposium on Romance Languages*. Edited by En K. Arregi, Z. Fagyal, S. Montrul and A. Tremblay. Amsterdam: John Benjamins, pp. 23–38.

Culberston, Jennifer, Hanna Jarvinen, Frances Heggarty, and Kenny Smith. 2019. Children's sensitivity to phonological and semantic cues during noun class learning evidence for a phonological bias. *Language* 95: 268–93. [CrossRef]

Cuza, Alejandro, and Roció Pérez-Tattam. 2016. Grammatical gender selection and phrasal word order in child heritage Spanish: A feature re-assembly approach. *Bilingualism: Language and Cognition* 19: 50–68. [CrossRef]

Cuza, Alejandro, Ana T. Pérez-Leroux, and Liliana Sánchez. 2013. The role of semantic transfer in clitic-drop among Chinese L1-Spanish L2 bilinguals. *Studies in Second Language Acquisition* 35: 93–125. [CrossRef]

D'Imperio, Mariapaola, Gorka Elordieta, Sónia Frota, Pilar Prieto, and Marina Viga'rio. 2005. Intonational phrasing in Romance: The role of syntactic and prosodic structure. In *Prosodies*. Edited by Sónia Frota, Marina Viga'rio and Maria J. Freitas. Berlin: Mouton de Gruyter, pp. 59–98.

Davidson, Lisa, and Daniel Erker. 2014. Hiatus resolution in American English: The case against glide insertion. *Language* 90: 482–514. [CrossRef]

Delattre, Pierre. 1965. *Comparing the Phonetic Features of English, French, German and Spanish: An Interim Report*. Philadelphia: Julius Groos Verlag.

Durvasula, Karthik, and Jimin Kahng. 2015a. Illusory vowels in perceptual epenthesis: The role of phonological alternations. *Phonology* 32: 385–416. [CrossRef]

Durvasula, Karthik, and Jimin Kahng. 2015b. *Phonological Alternations Modulate Illusory Vowels in Perceptual Epenthesis*. Tromso: Universitetet i Tromsoe.

Eichler, Nadine, Veronika Jansen, and Natascha Müller. 2013. Gender acquisition in bilingual children: French–German, Italian–German, Spanish–German and Italian–French. *International Journal of Bilingualism* 17: 550–72. [CrossRef]

Ettlinger, Marc, and Jennifer Zapf. 2011. The role of phonology in children's acquisition of the plural. *Language Acquisition* 18: 294–313. [CrossRef] [PubMed]

Fox, Robert A., James E. Flege, and Murray J. Munro. 1994. The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. *Journal of the Acoustic Society of America* 97: 2540–51. [CrossRef] [PubMed]

Frota, Sonia, Mariapaola D'Imperio, Gorka Elordieta, Pilar Prieto, and Marina Viga'rio. 2007. The phonetics and phonology of intonational phrasing in Romance. In *Prosodic and Segmental Issues in (Romance) Phonology*. Edited by Pilar Prieto, Joan Mascaro and Maria-Josep Sole. Amsterdam: John Benjamins, pp. 131–53.

Gathercole, Virginia. C. Mueller. 2002. Grammatical gender in bilingual and monolingual children: A Spanish morphosyntactic distinction. In *Language and Literacy in Bilingual Children*. Edited by D. Kimbrough Oller and Rebecca E. Eilers. Clevedon: Multilingual Matters, pp. 207–19.

Gilbert, Harvey R., Michael P. Robb, and Yang Chen. 1997. Formant frequency development: 15 to 36 months. *Journal of Voice* 11: 260–66. [CrossRef]

Gildersleeve-Neumann, Christina E., Elizabeth D. Pena, Barbara L. Davis, and Ellen S. Kester. 2009. Effects on L1 during early acquisition of L2: Speech changes in Spanish at first English contact. *Bilingualism: Language and Cognition* 12: 259–72. [CrossRef]

Goebel-Mahrle, Thomas, and Naomi L. Shin. 2020. A corpus study of child heritage speakers' Spanish gender agreement. *International Journal of Bilingualism* 2020: 1–17. [CrossRef]

Goldstein, Brian A., and Karen E. Pollock. 2000. Vowel errors in Spanish-speaking children with phonological disorders: A retrospective comparative study. *Clinical Linguistics and Phonetics* 14: 217–34.

Grinstead, John, Myriam Cantú-Sánchez, and Blanca Flores-Ávalos. 2008. Canonical and epenthetic plural marking in Spanish-speaking children with specific language impairment. *Language Acquisition* 15: 329–49. [CrossRef]

Gu, Chong. 2014. Smoothing Spline ANOVA Models: R Package Gss. *Journal of Statistical Software 58: 1–25.* Available online: http://www.jstatsoft.org/v58/i05/ (accessed on 1 July 2020).

Guasti, Maria. T., Anna Gavarro, Joke De Lange, and Claudia Caprin. 2008. Article omission across child languages. *Language Acquisition* 15: 89–119. [CrossRef]

Hawkins, Sarah. 2010. Phonological features, auditory objects, and illusions. *Journal of Phonetics* 38: 60–89. [CrossRef]

Haynes, Erin F., and Michael Taylor. 2014. An assessment of acoustic contrasts between short and long vowels using convex hulls. *Journal of the Acoustical Society of America* 136: 883–91. [CrossRef]

Hualde, José I. 2014. *Los sonidos del Espanõl*. Cambridge: Cambridge University Press.

Hualde, José I., Miquel Simonet, and Francisco Torreira. 2008. Postlexical contraction of nonhigh vowels in Spanish. *Lingua* 118: 1906–25. [CrossRef]

Hutchinson, Sandra P. 1974. Spanish vowel sandhi. In *Parasession on Natural Phonology of the Regional Meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society, pp. 752–62.

Johnson, Daniel E. 2015. Quantifying Overlap with Bhattacharyya's Affinity. *Paper presented at NWAV 44, Toronto, ON, Canada, October 25*. Available online: https://danielezrajohnson.shinyapps.io/nwav_44 (accessed on 28 August 2020).

Kehoe, Margaret M., and Conxita Lleó. 2017. Vowel reduction in German-Spanish bilinguals. In *Romance-Germanic Bilingual Phonology*. Edited by Mehmet Yavas, Margaret M. Kehoe and Walcir C. Cardoso. London: Equinox, pp. 14–37.

Kent, Raymond D., and Ann. D. Murray. 1982. Acoustic features of infant vocalic utterances at 3, 6 and 9 months. *Journal of the Acoustic Society of America* 72: 353–65. [CrossRef]

Kuchenbrandt, Imme. 2005. Gender acquisition in bilingual Spanish. In *Proceedings of the 4th International Symposium on Bilingualism*. Edited by James Cohen, Kara T. McAlister, Kellie Rolstad and Jeff MacSwan. Somerville: Cascadilla Press, pp. 1252–63.

Ladefoged, Peter. 2001. *Vowels and Consonants: An Introduction to the Sounds of Language*. Oxford: Blackwell.

Larranãga, Pilar, uijarro-Fuentes, and edro. 2012. Clitics in L1 bilingual acquisition. *First Language* 32: 151–75. [CrossRef]

Lleo', Conxita, and Katherine Demuth. 1999. Prosodic constraints on the emergence of grammatical morphemes: Crosslinguistic evidence from Germanic and romance languages. In *Proceedings of the 23rd annual Boston University Conference on Language Development*. Edited by Annabel Greenhill, Heather Littlefield and Cheryl Tano. Somerville: Cascadilla Press, pp. 407–18.

Lopez Ornat, Susana. 1997. What lies in between a pre-grammatical and a grammatical representation? *In Contemporary Perspectives on the Acquisition of Spanish, , Volume 1: Developing Grammars*. Edited by Ana T. Pérez-Leroux and William R. Glass. Somerville: Cascadilla Press, pp. 3–20

Lobanov, Boris M. 1971. Classification of Russian vowels spoken by different listeners. *Journal of the Acoustical Society of America* 49: 606–8. [CrossRef]

Mak Brian, Etienne Barnard. 1996. Phone clustering using the Bhattacharyya distance. In *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96*. Philadelphia: IEEE, vol. 4, pp. 2005–8.

Martínez-Gibson, Elizabeth. A. 2011. A comparative study on gender agreement errors in the spoken Spanish of heritage speakers and second language learners. *Porta Linguarum* 15: 177–93.

Mateu, Victoria E. 2015. Object Clitic Omission in Child Spanish: Evaluating Representational and Processing Accounts. *Language Acquisition* 22: 240–84. [CrossRef]

Mazzaro, Natalia, Laura Colantoni, and Alejandro Cuza. 2016. Age effects and the discrimination of consonantal and vocalic contrasts in heritage and native Spanish. In *Romance Linguistics 2013: Selected Proceedings of the 43th Linguistic Symposium on Romance Languages*. Edited by Christina Tortora, Marcel den Dikken, Ignacio L. Montoya and Teresa O'Neill. Amsterdam: John Benjamins, pp. 277–300.

McCloy, Daniel R. 2016. phonR: Tools for Phoneticians and Phonologists. R Package Version 1.0–7. Available online: http://drammock.github.io/phonR/ (accessed on 25 June 2020).

Menke, Mandy. 2010. Examination of the Spanish vowels produced by Spanish-English bilingual children. *Southwest Journal of Linguistics* 28: 98–135.

Miller, Karen, and Cristina Schmitt. 2010. Effects of variable input in the acquisition of plural in two dialects of Spanish. *Lingua* 120: 1178–93. [CrossRef]

Mohaghegh, Mercedeh. 2016. Connected Speech Processes and Lexical Access in Real-Time Comprehension. Unpublished Ph.D. dissertation, University of Toronto, Toronto, ON, Canada.

Montrul, Silvina. 2011. Morphological errors in Spanish second language learners and heritage speakers. *Studies in Second Language Acquisition* 33: 155–61. [CrossRef]

Montrul, Silvina, and Kim Potowski. 2007. Command of gender agreement in school-age Spanish-English bilingual children. *International Journal of Bilingualism* 11: 301–28. [CrossRef]

Montrul, Silvina, labakova, and oumyana. 2003. Competence similarities between native and near-native speakers: An investigation of the preterite/imperfect contrast in Spanish. *Studies in Second Language Acquisition* 25: 351–98. [CrossRef]

Montrul, Silvina, Rebecca Foote, and Silvia Perpiñán. 2008. Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language Learning* 58: 503–53. [CrossRef]

Montrul, Silvina, Justin Davidson, Israel de la Fuente, and Rebecca Foote. 2014. Early language experience facilitates the processing of gender agreement in Spanish heritage speakers. *Bilingualism: Language and Cognition* 17: 118–38. [CrossRef]

Morgan, Gareth P., M. Adelaida Restrepo, and Alejandra Auza. 2013. Comparison of Spanish morphology in monolingual and Spanish-English bilingual children with and without language impairment. *Bilingualism: Language and Cognition* 16: 578–96. [CrossRef]

Morrison, Geoffrey S. 2003. Perception and production of Spanish vowels by English speakers. In *Proceedings of the 15th International Congress of Phonetic Sciences*. Edited by Maria-Josep Sole', Daniel Recasens and Joaquin Romero. Adelaide: Causal Productions, pp. 1533–36.

Morrison, Geoffrey S. 2006. L1 and L2 Production and Perception of English and Spanish Vowels: A Statistical Modelling Approach. Unpublished Ph.D. dissertation, University of Alberta, Edmonton, AB, Canada.

Navarro Toma's, Tomás. 1970. *La pronunciacio'n del espanõl*. Madrid: CSIC.

Nearey, Terrance M. 1977. Phonetic Feature Systems for Vowels. Ph.D. dissertation, University of Alberta, Edmonton, AB, Canada.

Nevins, Andrew. 2014. Misheard lyrics Introduction: ABBA's Super Trouper. Available online: https://youtu.be/dBnhkwRmYuQ (accessed on 10 October 2020).

Nicoladis, Elena, and Kristan Marchak. 2011. Le carte blanc or la carte blanche? bilingual children's acquisition of french adjective agreement. *Language Learning* 61: 734–58. [CrossRef]

Ohala, John J. 1989. Sound change is drawn from a pool of synchronic variation. In *Language Change: Contributions to the Study of Its Causes*. Edited by Leiv Breivik and Ernst Jahr. New York: Mouton de Gruyter, pp. 173–98.

Ohala, John J. 1993. The phonetics of sound change. In *Historical Linguistics: Problems and Perspectives*. Edited by Charles Jones. London: Longman, pp. 237–78.

Pérez-Pereira, Miguel. 1991. The acquisition of gender: What Spanish children tell us. *Journal of Child Language* 18: 571–90. [CrossRef] [PubMed]

R Studio Team. 2020. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA. Available online: http://www.rstudio.com/ (accessed on 25 May 2020).

Repp, Bruno H. 1992. Perceptual Restoration of a "Missing" Speech Sound: Auditory Induction or Illusion? *Perception & Psychophysics 51: 14–32.* Available online: http://myaccess.library.utoronto.ca/login?qurl=https%3A%2F%2Fsearch.proquest.com%2Fdocview%2F58231001%3Faccountid%3D14771 (accessed on 28 August 2020).

Rogers, Henry. 2000. *The Sounds of Language: An Introduction to Phonetics*. London: Longman.

Rogers, Brandon M. 2012. Vowel Quality and Language Contact in Miami Cuban Spanish. Ph.D. dissertation, Brigham Young University, Provo, UT, USA.

Romanelli, Sofía, Andrea Menegotto, and Ron Smyth. 2018. Stress-induced acoustic variation in L2 and L1 Spanish vowels. *Phonetica* 75: 190–218. [CrossRef] [PubMed]

Ronquest, Rebecca. 2016. Stylistic variation in heritage Spanish vowel production. *Heritage Language Journal* 13: 275–96. [CrossRef]

Ronquest, Rebecca, and Rajiv Rao. 2018. Heritage Spanish phonetics and phonology. In *The Routledge Handbook of Spanish as a Heritage Language*. Edited by Kim Potowski and Javier Muñoz-Basols. Abingdon: Routledge, pp. 164–77.

Rvachew, Susan, Elzbieta. B. Slavinski, Megan Williams, and Carol L. Green. 1996. Formant frequencies of vowels produced by infants with and without early onset otitis media. *Canadian Acoustical Society* 24: 19–28.

Rvachew, Susan, Karen Mattock, Linda Polka, and Lucie Me'nard. 2006. Developmental and cross-linguistic variation in the infant vowel space: The case of Canadian English and Canadian French. *Journal of the Acoustic Society of America* 120: 2250–59. [CrossRef]

Schnitzer, Marc L., and Emily Krasinski. 1994. The development of segmental and phonological production in a bilingual child. *Journal of Child Language* 21: 585–622. [CrossRef]

Scontras, Gregory, Zuzanna Fuchs, and Maria Polinsky. 2015. Heritage language and linguistic theory. *Frontiers in Psychology* 6: 15–45. [CrossRef]

Scontras, Gregory, Maria Polinsky, and Zuzanna Fuchs. 2018. In support of representational economy: Agreement in heritage Spanish. *Glossa: A Journal of General Linguistics* 3: 1–28. [CrossRef]

Shoemaker, Ellenor, and Rebekah Rast. 2013. Extracting words from the speech stream at first exposure. *Second Language Research* 29: 165–83. [CrossRef]

Silva-Corvala'n, Carmen. 2014. *Bilingual Language Acquisition: Spanish and English in the First Six Years*. Cambridge: Cambridge University Press.

Snyder, William, Ann Senghas, and Kelly Inman. 2001. Agreement morphology and the acquisition of noun-drop in Spanish. *Language Acquisition* 9: 157–73. [CrossRef]

Song, Jae Y., Megha Sundara, and Katherine Demuth. 2009. Phonological constraints on children's production of English third person singular–s. *Journal of Speech, Language, and Hearing Research* 52: 623–64. [CrossRef]

Stoel-Gammon, Carol, and Karen Pollock. 2009. Vowel development and disorders. In *The handbook of Clinical Linguistics*. Edited by Martin J. Ball, Michael R. Perkins, Nicole Mu¨ller and Sara Howard. Hoboken: Blackwell, pp. 525–548.

Stoel-Gammon, Carol, and Anna Vogel Sosa. 2008. Phonological development. In *Language Development*. Edited by Erika Hoff and Marilyn Shatz. Oxford: Blackwell, pp. 238–56.

Strelluf, Christopher. 2016. Overlap among back vowels before /l/ in Kansas City. *Language Variation and Change* 28: 379–407. [CrossRef]

Vokic, Gabriela, and Jorge M. Guitart. 2009. On the realization of non-high Spanish vowels in contact within sequences of more than three vocoids: Spectrographic evidence and theoretical implications. *Southwest Journal of Linguistics* 28: 71–102.

Willis, Erik W. 2005. An initial examination of southwest Spanish vowels. *Southwest Journal of Linguistics* 24: 185–98.

Winter, Bodo. 2020. *Statistics for Linguists: An Introduction Using R*. New York and London: Routledge.

languages

MDPI

Article

(Divergent) Participation in the California Vowel Shift by Korean Americans in Southern California

Ji Young Kim [1,*] and Nicole Wong [2]

[1] Department of Spanish and Portuguese, University of California, Los Angeles, CA 90095, USA
[2] Independent Scholar, Santa Clara, CA 95050, USA; wong.nicolew@gmail.com
* Correspondence: jiyoungkim@ucla.edu

Received: 24 September 2020; Accepted: 2 November 2020; Published: 6 November 2020

Abstract: This study investigates the participation in the California Vowel Shift by Korean Americans in Los Angeles. Five groups of subjects participated in a picture narrative task: first-, 1.5-, and second-generation Korean Americans, Anglo-Californians, and (non-immigrant) Korean late learners of English. Results showed a clear distinction between early vs. late bilinguals; while the first-generation Korean Americans and the late learners showed apparent signs of Korean influence, the 1.5- and the second-generation Korean Americans participated in most patterns of the California Vowel Shift. However, divergence from the Anglo-Californians was observed in early bilinguals' speech. Similar to the late bilinguals, the 1.5-generation speakers did not systematically distinguish prenasal and non-prenasal /æ/. The second-generation speakers demonstrated a split-/æ/ system, but it was less pronounced than for the Anglo-Californians. These findings suggest that age of arrival has a strong effect on immigrant minority speakers' participation in local sound change. In the case of the second-generation Korean Americans, certain patterns of the California Vowel Shift were even more pronounced than for the Anglo-Californians (i.e., /ɪ/-lowering, / ɑ/-/ɔ/ merger, /ʊ/- and /ʌ/-fronting). Moreover, the entire vowel space of the second-generation Korean Americans, especially female speakers, was more fronted than that of the Anglo-Californians. These findings suggest that second-generation Korean Americans may be in a more advanced stage of the California Vowel Shift than Anglo-Californians or the California Vowel Shift is on a different trajectory for these speakers. Possible explanations in relation to second-generation Korean Americans' intersecting gender, ethnic, and racial identities, and suggestions for future research are discussed.

Keywords: Korean Americans; California Vowel Shift; second language phonology; bilingualism; immigrant minority speakers; sound change

1. Introduction

Over the past few decades, research on second language (L2) phonology has provided empirical evidence that early bilinguals are generally more successful in acquiring L2 speech sounds than late bilinguals (Flege et al. 1995, 1997; Flege and MacKay 2011; Stevens 1999; Yeni-Komshian et al. 2000). Models in L2 phonology, such as Flege (1995) Speech Learning Model (SLM) and Best and Tyler (2007) Perceptual Assimilation Model (PAM)-L2, posit that bilinguals' L1 and L2 phones interact in a common phonological space. Thus, the development of L2 sounds would depend on the perceptual similarity to existing L1 sounds. That is, bilinguals would assimilate an L2 sound to an L1 sound if the two are perceived identical or if the L2 sound is perceived as a deviant variant of the L1 sound. However, if an L2 sound is perceptually distinct from existing L1 sounds, bilinguals would create a new category. Early bilinguals tend to be successful at simultaneously maintaining language-internal and cross-linguistic contrasts (Chang et al. 2011) because they begin establishing L2 sounds when they are still in the process of acquiring language-general fine-grained acoustic features (Kuhl et al. 1992;

Werker and Tees 1984). For late bilinguals, on the other hand, L2 sounds are introduced to an already established L1 sound system. Thus, influence from L1 speech sounds would occur to a larger extent for late bilinguals than for early bilinguals.

In the case of immigrant populations, first-generation speakers (i.e., late bilinguals) are prone to having a foreign accent in the societal language despite long residence in the host country (Baker and Trofimovich 2005). With respect to children of immigrants who are early bilinguals of their home language and the societal language, the situation becomes complicated. Some speakers do not show any signs of foreign accent in the societal language (Lloyd-Smith et al. 2020), while others demonstrate phonetic features that are different from local mainstream varieties. For instance, immigrant minority speakers who acquired the societal language natively may use phonetic features that are present in their parents' foreign-accented speech, regardless of whether they speak their parents' language (Fought 2003; Mendoza-Denton 1999; Mendoza-Denton and Iwai 1993; Tsukada et al. 2005). Thus, while neurological maturation associated with age of acquisition plays an important role in L2 pronunciation, there are various extralinguistic factors other than age of acquisition that contribute to the development of L2 speech sounds (e.g., quantity/quality of L2 input, relative use of L1/L2, language attitude, identity, speech register) (Jia and Aaronson 2003; Flege 1999; Zampini 2008).

1.1. Ethnicity and Participation in Local Sound Change

Since the foundational work on African American English by Labov (1972), ethnicity has been considered as one of the key factors, along with age, gender, and social class, that condition language variation in a speech community (Boberg 2004). With regard to immigrant minority groups in North America with non-English speaking backgrounds, native-born children tend to display less ethnic identification than their foreign-born parents (Hoffman and Walker 2010; Weinfeld 1985, pp. 71–77). Thus, apart from producing more native-like speech sounds in English, they demonstrate stronger assimilation to local mainstream norms. Nevertheless, studies have shown that even those who were born and raised in North America and speak English natively demonstrate some speech patterns that are distinct from the local mainstream varieties. For instance, Casillas and Simonet (2016) examined the production (and perception) of the English low vowels /æ/ and /ɑ/ by two groups of Spanish-English sequential bilinguals residing in Southern Arizona: Mexican Americans born and raised in the US Southwest by Spanish-speaking parents from Northern Mexico (i.e., native-born) and late English learners born and raised in Spanish-speaking countries, who moved to the US Southwest and lived there around 10 years (i.e., foreign-born). Both groups acquired Spanish as their L1, but the native-born speakers became more dominant in English as they grew up to the point that they were no longer able to actively communicate in Spanish. Unlike Spanish which has only one low vowel /a/, Southern Arizona English has two low vowels, /ɑ/ and /æ/, the latter of which is lowered in non-prenasal contexts, similar to California English (see Section 1.2). Casillas and Simonet (2016) found that, although the bilinguals were able to distinguish the two vowel categories, the phonetic realizations of these vowels were different from the local mainstream norms. Both groups produced more fronted /ɑ/ than English monolinguals, assimilating to the Spanish central /a/. Regarding /æ/, the foreign-born speakers produced this vowel more back than the English monolinguals (i.e., assimilation to the Spanish central /a/), whereas the native-born speakers produced it higher (i.e., weaker /æ/-lowering). These findings suggest that, even after shift to English occurs, the speech of native-born speakers may diverge from local mainstream norms either by demonstrating patterns that are traceable to their heritage language or by participating to a lesser extent in the sound change of the local mainstream variety.

From a developmental point of view, it is important to note that immigrant minority speakers who no longer speak their heritage language or have only passive knowledge of it may use speech patterns that differentiate themselves from speakers of other ethnicities. According to Labov (2001, p. 506), "[a]ll speakers who are socially defined as white, mainstream, or Euro-American, are involved in [regional sound] changes to one degree or another." Certain patterns of regional sound change may also appear in the speech of some ethnic minority speakers that American society defines as "non-white" (e.g., Black, Hispanic, Native American, Asian). However, it is unlikely that ethnic minority speakers converge with Anglo-Americans in all aspects of their speech (Labov 2001, p. 507). Rather, they often take a different trajectory in regional sound change. For instance, studies have shown that US-born Latinos tend to resist prenasal /æ/-raising (i.e., split between prenasal and non-prenasal /æ/) and /u/-fronting, which are features that occur in many varieties of American English (Carter et al. 2020; Fought 1999; Roeder 2010; Thomas 2001). Carter et al. (2020) examined the speech of Miami-born Latinos whose parents immigrated from various Latin American countries and found that their English /u/ was more back than that of Anglo-Americans. Moreover, although the Latino speakers distinguished prenasal and non-prenasal /æ/, their /æ/ in both environments were more back than their Anglo counterparts. Resistance to prenasal /æ/-raising and /u/-fronting, which may be due to influence from the Spanish low central vowel /a/ and high back vowel /u/, has also been reported in the speech of Mexican Americans in other regions (California: Fought (1999, 2003), Michigan: Roeder (2010), Texas: Thomas (2000, 2001), Washington DC: (Tseng 2015)), although in some cases these patterns are conditioned by social factors such as gender, social class, and group affiliation (Fought 1999; Roeder 2010; Tseng 2015).

Compared to African American English and Chicano English, little research has been conducted on English spoken by Asian Americans. However, studies have found that Asian Americans, like other ethnic minority speakers, show a combination of resistance and assimilation to local sound change (Cheng 2016; Hall-Lew 2009; Hall-Lew and Starr 2010; Hoffman 2010; Ito 2010; Lee 2000, 2016). For instance, Hall-Lew (2009) found that Chinese Americans in San Francisco participated in back vowel fronting and low back vowel merger, which are two sound changes that characterize California English. However, at the same time, Chinese Americans in this region produce coda /l/-vocalization (e.g., pronouncing *cold* and *skill* as *code* and *skew*), which most likely is due to influence from Chinese phonology that lacks syllable-final /l/ (Hall-Lew and Starr 2010). This pattern appears even in speakers beyond the second generation who are English monolingual speakers. Ito (2010) examined the vowels of Hmong Americans in the Twin Cities area in Minnesota, and found that 1.5 generation and second-generation speakers were accommodating to the local /æ/-fronting, but distinguished the low back vowels /ɑ/ and /ɔ/ more clearly than Anglo-Americans who showed a trend toward near-merger.

With regard to Korean Americans, which is the target population of this study, Cheng (2016) demonstrated that Korean Americans in California participated in some aspects of the California Vowel Shift (e.g., /u/-fronting and /ɑ/-/ɔ/ merger) to the same degree as Anglo-Americans, while in others their patterns were either more pronounced (e.g., /ʊ/-fronting) or less pronounced (e.g., split between prenasal and non-prenasal /æ/) than the Anglo-Americans. Lee (2016) found that Korean Americans in Bergen County, New Jersey, which borders New York City, maintained the /ɑ/-/ɔ/ contrast and raised /ɔ/ in accordance with New York City English, but they did not produce the New York City English split-/æ/ system (Labov et al. 2006). While these studies did not discuss Korean Americans' divergence from the white regional norms as influence from Korean phonology, it is possible that their less pronounced split-/ae/ system is related to Korean vowels which do not demonstrate such patterns. Thus, it is important to examine Korean-accented English of late bilinguals, particularly that of first-generation Korean immigrants, to see whether Korean Americans' resistance to sound change in local mainstream varieties has to do with their exposure to Korean and Korean-accented English.

1.2. California Vowel Shift

California English is easily exposed to people in other regions through television and movies, and the speech styles of stereotypical Southern California personae portrayed in the media (e.g., Valley Girl and Surfer Dude) are often parodied (Pratt and D'Onofrio 2017). A good example of this are The Californians skits from NBC's late-night comedy show Saturday Night Live (SNL). Pratt and D'Onofrio (2017) analyzed the vowels produced by two characters in The Californians, and found that the actors talked with more open and protruded jaws and lips to comedically portray the Valley Girl and Surfer Dude personae. The use of such articulatory settings resulted in the production of lower and more retracted front vowels and more fronted back vowels when the actors played these characters than when they played non-Californian characters. Although without a doubt these performances are exaggerated, they reflect the vocalic changes that are underway in California, namely the California Vowel Shift.

Figure 1, created from data of millennial speakers reported in D'Onofrio et al. (2019)[1], demonstrates the vocalic changes involved in the California Vowel Shift. The California Vowel Shift is characterized by three main phenomena: (1) the low-back merger of /ɑ/ (e.g., *bot*) and /ɔ/ (e.g., *bought*), (2) the lowering and retraction of lax front vowels /ɪ/ (e.g., *bit*), /ɛ/ (e.g., *bet*), and /æ/ (e.g., *bat*), and (3) the fronting of high- and mid-back vowels /u/ (e.g., *boot*), /ʊ/ (e.g., *book*), /o/ (e.g., *boat*), and /ʌ/ (e.g., *but*) (D'Onofrio et al. 2016; D'Onofrio et al. 2019; Hagiwara 1997; Hall-Lew 2009; Hall-Lew et al. 2015; Hinton et al. 1987; Kennedy and Grama 2012; Podesva et al. 2015). Following the pattern of General American English presented in The Atlas of North American English (Labov et al. 2006), prenasal /æ/ in California English is tensed, resulting in a split between tensed /æ/ in a prenasal context (e.g., *ban*) and lowered /æ/ elsewhere (Eckert 2008). With regard to the back vowels /u/ and /o/, the fronting is more advanced after a coronal consonant (e.g., *too* and *toe*) due to its high F2 (i.e., fronted) environment and prohibited when followed by the velarized coda /-l/ (e.g., *cool* and *goal*), because of its low F2 (i.e., retracted) environment (Hall-Lew 2011).

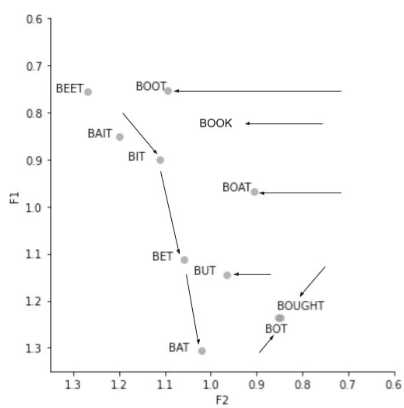


Figure 1. California Vowel Shift (adapted from D'Onofrio et al. (2019)).

While the California Vowel Shift has been understood as a chain shift affecting the front lax vowels /ɪ/, /ɛ/, and /æ/, the cause of the chain shift is under debate. Similar to the Canadian Vowel Shift in which /ɪ/, /ɛ/, and /æ/ are lowered due to the merging of /ɑ/ and /ɔ/ (Clarke et al. 1995), the lowering of /ɪ/, /ɛ/, and /æ/ in California English may also be the result of a pull-chain initiated by the /ɑ/-/ɔ/ merger. However, Kennedy and Grama (2012) found that some young California English speakers demonstrated the chain-shifted lowering of the front lax vowels, while maintaining /ɑ/ in the traditional low-central position in the vowel space. Moreover, while both male and female speakers exhibited similar F1 values for /ɪ/ and /ɛ/, the female speakers produced higher F1 values (i.e., lower vowel height) for /æ/ than the male speakers. Since women generally are leaders of linguistic change (Coates 1993; Labov 1990; Milroy and Milroy 1985; Trudgill 1972), the gender difference indicates that /æ/ is the most recent step of the chain shift (Kennedy and Grama 2012). Thus, Kennedy and Grama (2012) suggested an alternative explanation to the chain shift which involves a push-chain initiated by the

[1] Figure 1 was created based on the data of millennial speakers reported in Table A1 in D'Onofrio et al. (2019). Note that BOOK-type tokens (i.e., /ʊ/) were not examined in D'Onofrio et al. (2019), thus, we added the fronting of /ʊ/ in Figure 1 based on previous studies on the California Vowel Shift (e.g., Podesva et al. 2015; Pratt and D'Onofrio 2017).

lowering of /ɪ/, resulting in the lowering of /ɛ/ and subsequently the lowering of /æ/. This process is likely to be independent of the /ɑ/-/ɔ/ merger which in some cases occurs in the low-central position of /ɑ/ (Kennedy and Grama 2012) and in other cases is not fully instantiated (Hall-Lew 2009).

Chain shifts are claimed to occur in order to maintain enough phonetic distance between phonemes in the vowel space so that they are perceptually distinctive (D'Onofrio et al. 2019; Gordon 2011; Martinet 1952). If a phoneme moves within the vowel space, this leads to subsequent phonetic movements of neighboring vowels. Thus, in order to identify the vowel that triggered the movement, it is important to examine the temporal establishment of the chain shift in real or apparent time (D'Onofrio et al. 2019; Gordon 2011; Labov 2010, p. 145). That is, speakers from a certain age group in a more recent time period or younger speakers should exhibit more advanced movements of the chain shift than speakers of the same age group in an older time period or older speakers. D'Onofrio et al. (2019) conducted an apparent time study, comparing the vowels produced by speakers of four generations which were determined based on their birth year: Silent Generation (1928–1945), Baby Boomer (1946–1964), Generation X (1965–1980), and Millennial (after 1980). Results showed that, across the span of four generations, the speakers exhibited an overall reduction of dispersion mainly in the F2 dimension (i.e., frontedness), demonstrating a more advanced backing of front vowels and fronting of back vowels in younger generations. Most of these changes (i.e., /ɑ/-/ɔ/ merger, backing of /ɛ/ and /æ/, and fronting of postcoronal /u/ and /o/) appeared between the Silent and the Baby Boomer generations, suggesting that the horizontal compression of the vowel space occurred contemporaneously. In subsequent generations, continued /æ/-backing and /ɑ/-/ɔ/ merger were observed, as well as additional changes involving /ɪ/-backing, non-postcoronal /u/-fronting, lowering of /ɛ/, /æ/, and /i/, and raising of /ɑ/ and /ʌ/. These findings indicate that rather than a stepwise chain shift which has been previously claimed, the California Vowel Shift seems to show holistic compression of the vowel space. Phonologically speaking, this is contrary to the general tendency toward maximizing the phonetic space between phonemes as a means to maintain perceptual distinctiveness (Flemming 1996; Labov et al. 2006; Liljencrants and Lindblom 1972). Thus, D'Onofrio et al. (2019) proposed that the unexpected holistic compression at the root of the California Vowel Shift may be driven by speakers' projection of localized social meanings within a community (Eckert 1989; Fought 1999; Podesva 2011), not by purely phonological motivations. That is, it is possible that vowel space compression is achieved through speakers' manipulation of their articulatory settings (e.g., lowered jaw, protruded jaw and lips) (Pratt and D'Onofrio 2017) to index varied social meanings (e.g., young Californian, middle class membership, non-gang status, laid back, partier, urban, coastal) (D'Onofrio et al. 2019; Fought 1999; Podesva 2011; Podesva et al. 2015).

1.3. Vowels in Korean and Comparison between Korean and American English Vowel Systems

Modern South Korean has 7–8 monophthongs /i, e, (ɛ), a, ʌ, o, ɨ, u/ (Jang et al. 2015; Kang 2014; Kwak 2003; Lee 2000; Lee and Ramsey 2011; Yang 1996)[2]. Due to recent merger of the mid-front vowels /e/ and /ɛ/, which is most likely caused by the raising of /ɛ/, many Koreans no longer distinguish these vowels (Baker and Trofimovich 2005; Kang 2014; Kwak 2003; Jang et al. 2015; Lee and Ramsey 2011; Yang 1996). Studies examining Korean vowel change in apparent time (Jang et al. 2015; Kang 2014) have shown that the Korean /e/ and /ɛ/ are produced with overlapping F1 and F2 values across ages, except for some older speakers[3] who produced them distinctly, supporting that young-generation Koreans have a seven-vowel system with one mid-front vowel /e/ (Kwak 2003).

---

[2] In some cases, front rounded vowels /y/ and /ø/ may be additionally observed in the speech of older generation speakers, but in modern South Korean these sounds are mostly replaced by the diphthongs [we] and [wi], respectively (Ahn and Iverson 2007; Kwak 2003; Jang et al. 2015).

[3] Kang (2014) specified these speakers as male speakers born before 1962 (i.e., birth-year-based), while in Jang et al. (2015), these speakers were male and female speakers in their 60s (i.e., age-based). Since the data in Jang et al. (2015) were collected between 2014 and 2015, we speculate that these speakers were born between 1945 and 1955.

Figure 2, created from data of Korean speakers in their 20s in Kang and Kong (2016)[4], demonstrates the Korean vowel space. In the high region of the vowel space, Korean has three vowels /i/, /ɨ/, and /u/. While the Korean /i/ is acoustically similar to the corresponding English high front tense vowel /i/, the Korean /u/ is more back than the English high back tense vowel /u/ (Baker and Trofimovich 2005; Yang 1996; Yoon and Kim 2015). In fact, comparative studies of Seoul Korean and American English vowels have shown that the English /u/ is acoustically very similar to the Korean /ɨ/ (Baker and Trofimovich 2005; Yang 1996; Yoon and Kim 2015). The Korean /ɨ/, especially in the Seoul dialect, is undergoing change in progress in which younger-generation Koreans produce this vowel more fronted than older-generation Koreans (Jang et al. 2015; Kang 2014; Kang and Kong 2016; Lee et al. 2017)[5]. Thus, the overlap between the Korean /ɨ/ and the English /u/ may be due to the parallel fronting of the Korean /ɨ/ (Jang et al. 2015; Kang 2014; Kang and Kong 2016; Lee et al. 2017) and the English /u/ observed in most North American dialects (Labov et al. 2006). The Korean /u/ also exhibits fronting (Kang 2014; Lee et al. 2017), but not to the same extent as the Korean /ɨ/ and the English /u/.

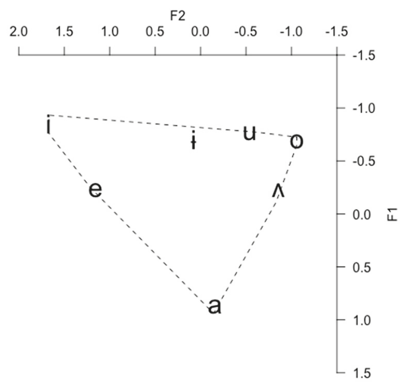


Figure 2. Korean vowel space of young generation Koreans (adapted from Kang and Kong (2016)).

In the mid region of the vowel space, Korean has two vowels /e/ and /o/ (or three vowels for speakers who do not exhibit the /e/-/ɛ/ merger). Unlike the English /e/ and /o/ which are slightly diphthongized (i.e., [eɪ], [oʊ]), the Korean /e/ and /o/ are purely monophthongal. In the case of the Korean /e/, it is also acoustically distinct from the other English mid-front vowel /ɛ/, in that the Korean /e/ is positioned higher and more fronted in the vowel space than the English /ɛ/ (Baker and Trofimovich 2005). Rather than the English /ɛ/, the Korean /e/ is acoustically more similar to the English front lax vowel /ɪ/ (Baker and Trofimovich 2005). Both the English /ɛ/ and /ɪ/ are experiencing lowering and retraction in General American English, except for the South (Labov et al. 2006). As for the Korean /o/, this vowel is positioned very high in the vowel space, close to the Korean /u/ (see Figure 2). Studies focusing on Seoul Korean have shown that the main distinction between the Korean /o/ and /u/ is in the front-back dimension (F2); the Korean /u/ is more fronted than the Korean /o/ (Jang et al. 2015; Kang and Kong 2016; Lee et al. 2017; Yoon and Kim 2015). These patterns were more clearly demonstrated in the speech of younger generation Koreans than older generation Koreans (Kang 2014; Kang and Kong 2016; Lee et al. 2017[6]). Thus, the vowel change in Seoul Korean can be explained through a chain shift initiated by /o/-raising,

[4] Figure 2 was created by averaging the data of male and female speakers in their 20s reported in Table 1 in Kang and Kong (2016).

[5] In a cross-dialectal study, Lee et al. (2017) found that speakers of other Korean dialects (i.e., South Jeolla, South Gyeongsang, and Jaeju) showed converging patterns to Seoul Korean.

[6] Despite dialectal differences, Lee et al. (2017) showed that the /o/-raising in other Korean dialects demonstrates converging patterns to the Seoul dialect, similar to the case of the fronting of /ɨ/ and /u/.

which led to the fronting of /u/ and /ɨ/ (Kang 2014; Kang and Kong 2016; Lee et al. 2017). In the low region of the vowel space, Korean has one low-central vowel /a/ which is more fronted than the corresponding English low back vowel /ɑ/ (Sohn 1999; Yang 1996). With regard to the Korean near-low back vowel /ʌ/, it is more back than the corresponding English /ʌ/ (Yang 1996) which is centralized [ɐ] in General American English, except for the Inland North (Labov et al. 2006).

Yang (1996) explained the cross-linguistic differences between Korean and English vowels through Lindblom's theory of adaptive dispersion (Lindblom 1990; Lindblom and Engstrand 1989), which proposes that speakers control sufficient perceptual contrast between phonemes, while monitoring a tradeoff between articulatory economy and perceptual distinctiveness. Yang (1996) argued that the English vowel space is characterized by the vertical expansion of low vowels resembling a rectangle, whereas the Korean vowel space is characterized by the horizontal expansion of high vowels resembling a triangle (Yang 1996). Using Lindblom (1990, p. 21) formula of perceptual distance, Yang (1996) demonstrated that the distance between the two extreme high vowels /i/ and /u/ was larger for Korean than for English. Given that Korean has more vowels in the high region of the vowel space (i.e., /i, ɨ, u/) than English (i.e., /i, u/), English-like fronting of /u/ would be restricted for the Korean /u/, since this would lead to an overlap with the Korean /ɨ/, causing perceptual confusion. The English /u/, on the other hand, is free to move forward without encroaching upon the space of other vowels (Yang 1996). With regard to the low region, Korean only has one vowel /a/, while English has two vowels /æ/ and /ɑ/. Thus, the Korean /a/ can be placed in the middle in the front-back dimension (i.e., corner of a regular triangle) without crowding into the space of other vowels, whereas the English /æ/ and /ɑ/ should be placed apart to maintain sufficient space between them. Yang (1996) also found that the perceptual distance between /i/ and /e/ and between /u/ and /o/ was larger for English than for Korean, which is in line with the predictions of the theory of adaptive dispersion (Lindblom 1990; Lindblom and Engstrand 1989). That is, the larger distance between the high and mid vowels in English is likely to be linked to English having intervening lax vowels /ɪ/ and /ʊ/, while Korean does not (Yang 1996).

Since Korean does not have tense-lax vowel contrasts like the English /i/-/ɪ/ and /u/-/ʊ/, Korean speakers often demonstrate difficulty in distinguishing these vowels (Baker and Trofimovich 2005; Flege et al. 1997). Baker and Trofimovich (2005) examined the acoustic properties of Korean and English vowels produced by early and late Korean-English bilinguals with varying length of residence in the US (i.e., 1 year and 7 years). They found that the late bilinguals, regardless of their length of residence in the US, produced the English /i/ and /ɪ/ and the English /u/ and /ʊ/ as two single categories which acoustically overlapped with the Korean /i/ (=English /i, ɪ/) and the Korean /u/ or /ɨ/ (=English /u, ʊ/). The late bilinguals also produced the English /ɛ/ and /æ/ as a single category, but they were dissimilar from the Korean /e/ (merged with the Korean /ɛ/). That is, the late bilinguals assimilated the English /i, ɪ/ to the Korean /i/ and assimilated the English /u, ʊ/ to the Korean /u/, while they created a new vowel category for the merged English /ɛ/-/æ/ (Baker and Trofimovich 2005). As for the early bilinguals, Baker and Trofimovich (2005) found that, regardless of the length of residence, they produced the English /i/ and /ɪ/ distinctly and their English /i/ overlapped with their Korean /i/. As for the other vowel contrasts, the early bilinguals who resided in the US for one year demonstrated similar patterns as the late bilinguals (i.e., merged English /u/-/ʊ/ and merged English /ɛ/-/æ/), whereas those with longer length of residence successfully distinguished these contrasts. For the latter group, the English /u/ overlapped with the Korean /u/, but the English /ʊ/ was produced as a separate category. Both early bilingual groups produced the English /ɪ/ as the Korean /e/. Thus, while the adult bilinguals who had three categories for the six English vowels /i, ɪ, u, ʊ, ɛ, æ/ (i.e., merged /i/-/ɪ/, merged /u/-/ʊ/, and merged /ɛ/-/æ/), the recently-arrived early bilinguals had four categories (i.e., /i/, /ɪ/, merged /u/-/ʊ/, and merged /ɛ/-/æ/), and the early bilinguals with longer length of residence distinguished all six vowels.

2. The Present Study

The goal of this study is to examine whether Korean Americans in Los Angeles participate in the California Vowel Shift. We focus on four vowel changes involved in the California Vowel Shift: (1) the lowering and retraction of front lax vowels /ɪ/, /ɛ/, and /æ/, (2) the split between non-prenasal /æ/ and prenasal /æN/, (3) the merging of low back vowels /ɑ/-/ɔ/, and (4) the fronting of /u/, /ʊ/, and /ʌ/. If Korean Americans do not exhibit the above-mentioned patterns, we explore whether their resistance to local sound change can be explained through influence from Korean phonology. To better understand this, we compare Korean Americans of three generations (first-generation, 1.5-generation, and second-generation) with Anglo-Californians and Korean international students who are late bilinguals. We predicted that, due to age effects, first-generation speakers (i.e., late bilinguals) would demonstrate stronger influence from Korean phonology than 1.5- and second-generation speakers (i.e., early bilinguals) and, thus, participate less in the California Vowel Shift. Regarding the early bilinguals, influence from Korean phonology, if any, would appear to a lesser extent for the second-generation speakers than for the 1.5-generation speakers.

With respect to patterns reflecting influence from Korean phonology, we base our predictions on the findings of Korean-English late bilinguals (Baker and Trofimovich 2005) and cross-linguistic differences between Korean and English vowels (Baker and Trofimovich 2005; Yang 1996; Yoon and Kim 2015). Regarding the lowering and retraction of /ɪ/, we predicted that, if influence from Korean phonology occurs, Korean Americans would merge this vowel with /i/, thus they would not participate in the lowering and retraction of /ɪ/. Moreover, Korean Americans would create a new single /ɛ/-/æ/ category in which /æ/ merges with /ɛ/ (Baker and Trofimovich 2005). Thus, when examining these vowels separately, Korean Americans' /ɛ/ may demonstrate lowering and retraction, but their /æ/ would not, because it would be positioned higher in the vowel space due to the merger with /ɛ/. Additionally, in the case of the English /æ/, Korean Americans would not exhibit a split-/æ/ system, because Korean does not have an equivalent phonological pattern. With regard to the /ɑ/-/ɔ/ merger, no Korean vowel acoustically overlaps with any of these vowels. The closest vowels are the Korean /a/ (low central) and /ʌ/ (near-low back) (Baker et al. 2002; Trofimovich et al. 2011; Tsukada et al. 2005). Thus, we predicted that, if influence from Korean occurs, they would either assimilate both the English /ɑ/ and the English /ɔ/ to the Korean /a/ (Outcome 1: Participation in /ɑ/-/ɔ/ merger but more fronted than expected) or distinguish them by assimilating the English /ɑ/ to the Korean /a/ and assimilating the English /ɔ/ to the Korean /ʌ/ (Outcome 2: No participation in /ɑ/-/ɔ/ merger). As for the fronting of /u/ and /ʊ/, Korean Americans would not participate or participate less in the fronting of these vowels, because they would assimilate both vowels to the Korean /u/ (Baker and Trofimovich 2005) which is fronted, but not to the same extent as in English (Kang 2014; Yang 1996). Similarly, Korean Americans' English /ʌ/ would be produced more back due to influence from the Korean /ʌ/ (Yang 1996).

2.1. Participants

In total, 37 Korean Americans, 4 Korean international students, and 5 Anglo-Americans participated in the study. The Korean Americans were residents of Los Angeles County in Southern California and consisted of three immigrant generations: first-generation (GEN1), 1.5-generation (GEN1.5), and second-generation (GEN2). The language background of each group is presented in Table 1. The GEN1 group (N = 8; 4F, 4M) were Koreans born and raised in South Korea (Seoul: 6, Daegu: 1, Yeongju: 1[7]) who immigrated to the US as adults (25.8 years). They had spent an average of 26.8 years in California at the time of data collection and spoke both English and Korean on a daily basis (English: 53.7%, Korean: 46.3%). They reported that they learned English (L2) during middle school or high school (12.8 years) and rated their English intermediate-level proficiency (2.7) on a 5-point Likert

[7] Both Daegu and Yeongju are cities in North Gyeongsang region in South Korea.

scale (5 = native). The GEN1.5 group (N = 4; 1F, 3M) were also Koreans born in South Korea (Seoul: 1, unspecified: 3), but they immigrated to the US in late childhood (10.5 years). As the GEN1 group, the GEN1.5 speakers learned English as an L2 and lived in the US for a long period of time (13.3 years). However, they used English (90%) much more frequently than Korean (10%), and rated their English (5 out of 5) as proficient as or more proficient than their native Korean (4.3). The GEN2 group (N = 25; 15F, 10M) were Koreans who were either born in Los Angeles County (N = 19) or born in South Korea (N = 6; Seoul: 4, Daegu: 1, unspecified: 1) and moved to Los Angeles County at age 3 or younger. All of their parents were first-generation immigrants from Korea (Seoul and Gyeonggi: 13, Gyeongsang: 2, Jeolla: 1, Chungcheong: 1, mixed: 2, unspecified: 6[8]). Seventeen of them were Korean-English early sequential bilinguals, while eight speakers acquired both languages simultaneously. Similar to the GEN1.5 group, the GEN2 speakers used English most of the time (English: 83.2%, Korean: 16.8%), but unlike the GEN1.5 group they rated their English proficiency (5 out of 5) much higher than their Korean proficiency (2.9 out of 5). Both the GEN1.5 and the GEN2 groups reported that, growing up, Korean was the main language of communication at home.

In this study we also included 4 Korean international students (KOR) as a baseline for Korean-accented English. The KOR speakers were born and raised in Seoul, South Korea, and came to the US to complete their undergraduate or graduate studies. These speakers were very similar to the GEN1 speakers in that they arrived in the US as adults (22.8 years), were late L2 learners of English (12.5 years), and used both English and Korean on a daily basis (English: 47.5%, Korean: 52.5%). However, compared to the GEN1 speakers, they spent less time in the US (4 years). Thus, if phonetic transfer from L1 Korean to L2 English appears, it would be strongest in this group. Lastly, 5 Anglo-Americans (2F, 3M) participated in this study as a control group for California English. All of these speakers were born and raised in Southern California (2 in Los Angeles County, 3 in San Diego County).

All participants read and signed a written informed consent form before they participated in the study. The study was conducted in accordance with the Declaration of Helsinki and the protocol was approved by the School of Literatures, Cultures, and Linguistics Institutional Review Board (SLCL-IRB) of the University of Illinois at Urbana-Champaign.

Table 1. Participants' language backgrounds (standard deviation is presented in parentheses).

	GEN1 (N = 8)	GEN1.5 (N = 4)	GEN2 (N = 25)	KOR (N = 4)	CA (N = 5)
Gender	4F, 4M	1F, 3M	15F, 10M	2F, 2M	2F, 3M
Age (years)	49.5 (9.3)	23.8 (3.8)	24.7 (4.4)	26.8 (6.1)	30.8 (15.4)
Age of arrival (years)	25.8 (5)	10.5 (1.7)	0.5 (0.9)	22.8 (7.1)	-
Length of residence (years)	26.8 (10.6)	13.3 (4.9)	-	4 (4.3)	-
English use	53.7% (24.5)	90% (4)	83.2% (15.8)	47.5% (5)	82% (18.3)
Korean use	46.3% (2.4)	10% (4)	16.8% (15.8)	52.5% (5)	-
English proficiency 1–5 (=Native)	2.7 (0.8)	5 (0)	5 (0.1)	3.7 (0.6)	5 (0)
Korean proficiency 1–5 (=Native)	5 (0)	4.3 (1)	2.9 (0.7)	5 (0)	-

2.2. Data Collection and Analysis

Participants' English speech data were collected during the spring and summer of 2012. In order to elicit different speech styles, we conducted a reading task (i.e., controlled speech) and a picture description task (i.e., narrative speech). For the reading task the participants read out loud a passage from Aesop's fables *The North Wind and the Sun* and for the picture description task they narrated the story of a wordless picture book *Frog, Where are You?* (Mayer 1969). In this study we will only

8 The areas are divided by regions, indicated by the suffix *Do* in Korean, in which different varieties of Korean are spoken.

report our findings from the narrative speech. Speech productions were audio-recorded using an AKG C520 head-mounted microphone and a Zoom H4n digital recorder with a sampling rate of 44.1 kHz and a sample size of 16 bits. The recordings were conducted in a quiet enclosed space in various locations in Los Angeles County (e.g., participants' home, furnished room in a church). In the case of two KOR speakers and two CA speakers, the recordings were conducted in a sound-attenuated booth at a public university in Illinois (KOR: 2, CA: 1)[9] and in Arizona (CA: 1)[10]. After completing the tasks, the participants filled out a language background questionnaire.

Participants' speech was orthographically transcribed on Praat TextGrids (Boersma and Weenink 2020) and forced alignment was performed using the Montreal Forced Aligner (McAuliffe et al. 2017), which generated a word tier and a phone tier. We extracted the F1 and the F2 values at the midpoint of 9 English monophthongs /i, ɪ, ɛ, æ, ɑ, ɔ, ʌ, ʊ, u/ and the durations of these vowels using a Praat script. For convenience purposes, we classified the vowels using the ARPAbet symbols: IY (=/i/), IH (=/ɪ/), EH (=/ɛ/), AE (=/æ/), AA (=/ɑ/), AO (=/ɔ/), AH (=/ʌ/), UH (=/ʊ/), and UW (=/u/). For AE, we further divided them into AE and AEN based on whether they preceded a non-nasal or a nasal consonant. A total of 28,948 tokens were obtained. In this study, we only considered vowels with primary stress and excluded vowels produced in fillers (e.g., *um*) or monosyllabic function words (e.g., *in*). Any tokens that were misaligned, too short in duration (<50 ms), or were produced with a creak, laughter, or background noise were excluded from the analyses. Moreover, in order to ensure reliable boundaries between the vowels and their neighboring sounds, we additionally excluded tokens following vowels, glides, or /r/, or tokens preceding vowels, glides, or liquids (Podesva et al. 2015). After this process, 2690 tokens remained (IY: 357, IH: 302, EH: 410, AE(N): 319, AA: 202, AO: 439, AH: 289, UH: 247, UW: 125). Raw F1 and F2 values (Hz) of these tokens were converted to a bark scale (Traunmüller 1997) and then normalized using Lobanov (1971) z-score procedure in the *phonR* package (McCloy 2016) in R (R Development Core Team 2020).

All statistical analyses and data visualizations were conducted using R (R Development Core Team 2020). In this study we examined the following patterns involved in the California Vowel Shift: (1) lowering and retraction of IH, EH, and AE, (2) AE-AEN split, (3) AA-AO merger, and (4) fronting of UW, UH, and AH. For the first three patterns (i.e., lowering/retraction, split, and merger), we analyzed both vowel height (F1) and frontedness (F2) and for the last pattern (i.e., fronting), we only analyzed vowel frontedness (F2). For statistical analyses, we performed linear mixed effects modeling in the *lme4* package (Bates et al. 2015). All fixed effects were contrast coded using simple coding in which each level is compared to the reference level and the intercept is the grand mean. The fixed and random effects of the model used in each pattern are presented in the following section. The best fitting model was selected through backward elimination and model comparisons were done with likelihood ratio tests using the *anova()* function. Although examining variation in gender was not the main purpose of the study, we included gender as a fixed effect in all models, due to its important role in sound change (i.e., female speakers are generally the leaders of linguistic change) (Coates 1993; Labov 1990; Milroy and Milroy 1985; Trudgill 1972). The p-values were obtained via Satterthwaite approximation in the *lmerTest* package (Kuznetsova Alexandra and Rune H. B. 2017). When significant interactions were found, we conducted post-hoc pairwise comparisons in the *emmeans* package (Lenth 2020).

3. Results

3.1. Phonemic Status of English Vowels

Figure 3 demonstrates the vowel space by group and gender based on their mean normalized F1 and F2 values of each vowel. Before looking into the patterns of the California Vowel Shift, we first

---

9 The two KOR speakers and the CA speaker had spent 2 years, 2 years, and 1 year, respectively, in Illinois at the time of data collection.

10 The CA speaker was temporarily visiting Arizona at the time of data collection, but was residing in Los Angeles County.

examined the phonemic status of the vowels in each group and checked whether the Korean Americans distinguished the vowel pairs IY-IH, EH-AE, UW-UH, and AH-AO. We analyzed the normalized F1 and F2 values to compare the height and the frontedness of the vowel pairs. We performed linear mixed effects modeling with vowel and gender as fixed effects and participant and item as random effects. Table 2 summarizes the statistical significance of the vowel contrasts in each group.

Table 2. Summary of the statistical results of English vowel contrasts.

	IY vs. IH	EH vs. AE	UH vs. UW	AH vs. AO
CA	F1: ***/F2: ***	F1: ***/F2: **	F1: ***/F2: n.s.	F1: ***/F2: **
GEN2	F1: ***/F2: ***	F1: ***/F2: ***	F1: ***/F2: n.s.	F1: ***/F2: ***
GEN1.5	F1: ***/F2: ***	F1: ***/F2: ***	F1: **/F2: n.s.	F1: */F2: $p = 0.057$
GEN1	F1: n.s./F2: *	F1: ***/F2: n.s.	F1: n.s./F2: n.s.	F1: n.s./F2: n.s.
KOR	F1: n.s./F2: *	F1: */F2: n.s.	F1: n.s./F2: n.s.	F1: n.s./F2: n.s.

***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, n.s.: non-significant.

There was a clear distinction between the GEN2 and the GEN1.5 speakers (i.e., early bilinguals) and the GEN1 speakers (i.e., late bilinguals), in that the former groups patterned like the CA speakers and the latter group patterned like the KOR speakers. The CA speakers and the early bilinguals successfully distinguished all four vowel contrasts using vowel height and frontedness except in the case of the UH-UW contrast which was distinguished by just vowel frontedness. In comparison, the late bilinguals were able to distinguish only the front vowel contrasts either using vowel height (i.e., EH-AE) or vowel frontedness (i.e., IY-IH), while failing to distinguish the back vowel contrasts UH-UW and AH-AO.

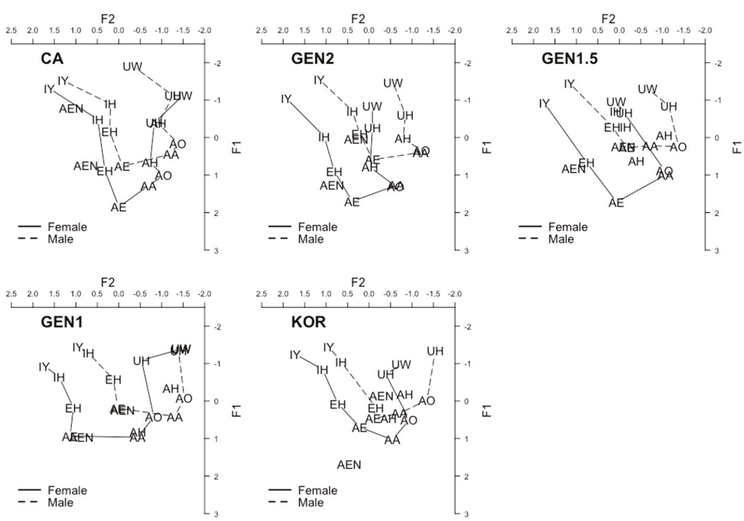


Figure 3. Normalized F1-F2 space by group and gender.

3.2. *Lowering and Retraction of IH, EH, and AE*

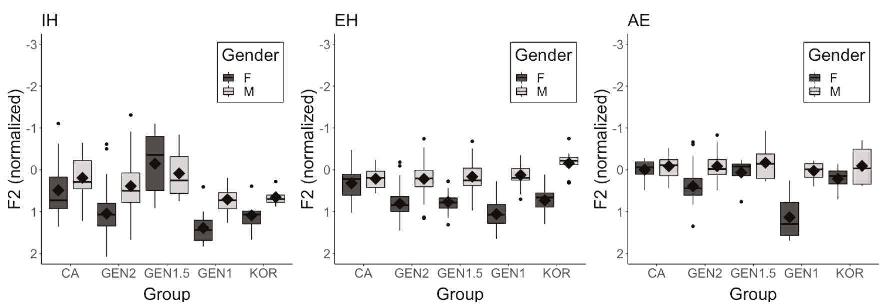
For the lowering and retraction of front lax vowels IH, EH, and AE, we examined the effects of group and gender on the F1 (vowel height) and the F2 (vowel frontedness) of each of these vowels. We performed linear mixed effects modeling with group (CA, GEN2, GEN1.5, GEN1, KOR), gender (female, male), and the interaction between group and gender as fixed effects and participant and item as random effects. For IH, we additionally included the following segment (nasal, non-nasal) as a fixed effect, given that studies have shown that this vowel becomes raised to IY before a nasal consonant (e.g., *thing*) (Hinton et al. 1987). We also included the nasality of the following segment as a fixed effect for the analysis of EH. As for AE, only tokens preceding a non-nasal consonant were examined. Further analysis of the effect of the nasality of the following segment on the realization of AE (i.e., AE-AEN split) will be presented in Section 3.3. The best-fitting models included random intercepts for subject and item without any random slope, except for the F1 of IH, which included a by-item random slope for gender.

Results of IH showed that there was a main effect of group (GEN2) on both the F1 ($\beta = 0.375$, SE = 0.151, t = 2.48, $p < 0.01$) and the F2 ($\beta = 0.269$, SE = 0.117, t = 2.307, $p < 0.05$), suggesting that the GEN2 speakers produced IH significantly lower[11] and more fronted than the CA speakers (i.e., reference level). The GEN1 and the KOR speakers also produced this vowel significantly more fronted than the CA speaker (GEN1: $\beta = 0.76$, SE = 0.137, t = 5.551, $p < 0.001$; KOR: $\beta = 0.587$, SE = 0.167, t = 3.509, $p < 0.001$). With regard to the effect of gender, overall, the female speakers (i.e., reference level) produced IH significantly lower and more fronted than the male speakers (F1: $\beta = -0.46$, SE = 0.128, t = −3.582, $p < 0.001$; F2: $\beta = -0.497$, SE = 0.093, t = −5.343, $p < 0.001$). For the F2, we found a significant interaction between group (GEN2) and gender ($\beta = -0.548$, SE = 0.233, t = −2.355, $p < 0.05$) and an interaction approaching significance between group (GEN1) and gender ($\beta = -0.504$, SE = 0.272, t = −1.852, $p = 0.071$). That is, the gender difference was larger for the GEN2 and the GEN1 speakers than for the CA speakers. Figure 4 presents the normalized F2 of IH, EH, and AE across groups and genders. Higher F2 values indicate more fronted realizations. Post-hoc pairwise comparison results confirmed that, while the CA speakers did not show significant difference between female and male speakers, the female GEN2 speakers and the female GEN1 speakers produced IH significantly more fronted than their male counterparts (GEN2: $\beta = 0.814$, SE = 0.09, t = 9.079, $p < 0.001$; GEN1: $\beta = 0.769$, SE = 0.165, t = 4.655, $p < 0.01$). Among the female speakers, the GEN1 speakers and the KOR speakers demonstrated the most fronted realizations of IH. When comparing across groups, the GEN1 speakers produced significantly more fronted IH than all the other groups, except for the KOR speakers (GEN1 vs. CA: $\beta = -1.012$, SE = 0.186, t = −5.435, $p < 0.001$; GEN1 vs. GEN2: $\beta = -0.469$, SE = 0.128, t = −3.65, $p < 0.05$; GEN1 vs. GEN1.5: $\beta = -1.226$, SE = 0.22, t = −5.584, $p < 0.001$). The KOR speakers also produced (marginally) significantly more fronted IH compared to the CA speakers ($\beta = -0.762$, SE = 0.236, t = −3.23, $p = 0.064$) and the GEN1.5 speakers ($\beta = -0.976$, SE = 0.261, t = −3.745, $p < 0.05$). Moreover, the GEN2 speakers produced significantly more fronted IH than the GEN1.5 speakers ($\beta = 0.758$, SE = 0.195, t = 3.891, $p < 0.05$) and the CA speakers ($\beta = -0.544$, SE = 0.155, t = −3.502, $p < 0.05$). The F2 of the GEN1.5 and the CA speakers did not differ. Therefore, female speakers' IH frontedness can be summarized with the following order: KOR, GEN1 > GEN2 > GEN1.5, CA. No group difference was found among the male speakers, except for the difference between the GEN1 and the GEN2 speakers, in which the former produced significantly more fronted IH than the latter ($\beta = -0.513$, SE = 0.14, t = −3.671, $p < 0.05$). Regarding the effect of the following segment, we found that overall, IH preceding a nasal consonant (i.e., reference level) were significantly more fronted than in other contexts ($\beta = 0.183$, SE = 0.084, t = 2.169, $p < 0.05$), but they did not differ in height.

With regard to EH, there was a main effect of group (GEN1) on both the F1 ($\beta = -0.397$, SE = 0.172, t = −2.311, $p < 0.05$) and the F2 ($\beta = 0.306$, SE = 0.11, t = 2.796, $p < 0.01$), which suggests that the GEN1 speakers produced EH significantly higher and more fronted than the CA speakers (i.e., reference level). The GEN2 speakers also produced EH significantly more fronted than the CA speakers (GEN2: $\beta = 0.249$, SE = 0.09, t = 2.777, $p < 0.05$), but the vowel height did not differ between the two groups. As found above, the female speakers (i.e., reference level) produced EH lower and more fronted than the male speakers (F1: $\beta = -0.777$, SE = 0.116, t = −6.723, $p < 0.001$; F2: $\beta = -0.605$, SE = 0.074, t = −8.164, $p < 0.001$). Moreover, we found a significant interaction between gender and

[11] In fact, among the five groups, GEN2 speakers' IH was produced with the lowest vowel height. Results of the same model with the GEN2 speakers as the reference group (instead of the CA speakers) confirmed that, except for the GEN1.5 speakers, the GEN2 speakers produced IH significantly lower than the other groups (CA: $\beta = -0.375$, SE = 0.151, t = −2.48, $p < 0.05$; GEN1: $\beta = -0.699$, SE = 0.124, t = −5.622, $p < 0.001$; KOR: $\beta = -0.603$, SE = 0.17, t = −3.556, $p < 0.001$).

group (KOR) for the F1 (β = 0.932, SE = 0.406, t = 2.295, $p < 0.05$), which indicates that the gender difference was larger for the KOR speakers than for the CA speakers. Post-hoc pairwise comparison results confirmed that female and male KOR speakers' EH did not differ in vowel height, whereas the female CA speakers produced this vowel significantly lower than the male CA speakers (β = 1.034, SE = 0.258, t = 4.005, $p < 0.05$). (Marginally) Significant gender differences were also found in the GEN2 (β = 1.017, SE = 0.117, t = 8.727, $p < 0.001$) and the GEN1 speakers (β = 0.737, SE = 0.223, t = 3.302, $p = 0.054$). The post-hoc test results also showed that the female GEN2 speakers produced EH with similar vowel height as the female CA and the female GEN1.5 speakers, whereas they produced EH significantly lower than the female GEN1 (β = 0.639, SE = 0.179, t = 3.577, $p < 0.05$) and the female KOR speakers (β = 0.774, SE = 0.23, t = 3.362, $p < 0.05$). No group difference in vowel height was found among the male speakers. For the F2, we found significant interactions between gender and all Korean groups, except for the GEN1.5 speakers (GEN2: β = −0.413, SE = 0.179, t = −2.308, $p < 0.05$; GEN1: β = −0.641, SE = 0.217, t = −2.954, $p < 0.01$; KOR: β = −0.533, SE = 0.261, t = −2.045, $p < 0.05$). This suggests that the gender difference was larger for these groups than for the CA speakers. Post-hoc pairwise comparison results confirmed that the female speakers in these groups produced EH significantly more fronted than the male speakers (GEN2: β = 0.623, SE = 0.074, t = 8.431, $p < 0.001$; GEN1: β = 0.851, SE = 0.143, t = 5.945, $p < 0.001$; KOR: β = 0.743, SE = 0.206, t = 3.6, $p < 0.05$), whereas no gender difference was found for the CA speakers. The post-hoc test results also revealed that the female CA speakers produced EH significantly less fronted than the female GEN2 (β = −0.456, SE = 0.131, t = −3.489, $p < 0.05$) and the female GEN1 speakers (β = −0.627, SE = 0.161, t = −3.888, $p < 0.05$). No group difference in F2 was found for the male speakers. Lastly, there was no main effect of the following segment on either measures.



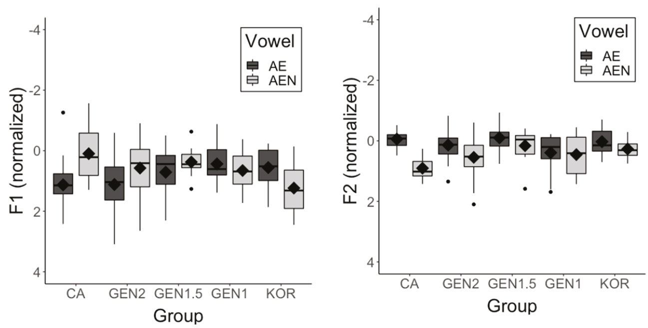
(**a**) (**b**) (**c**)

Figure 4. Normalized F2 across groups: (**a**) IH; (**b**) EH; (**c**) AE.

Regarding (non-prenasal) AE, similar to the case of EH, we found a main effect of group (GEN1) on both the F1 (β = −0.613, SE = 0.293, t = −2.092, $p < 0.05$) and the F2 (β = 0.557, SE = 0.183, t = 3.049, $p < 0.01$), indicating that the GEN1 speakers produced AE significantly higher and more fronted than the CA speakers (i.e., reference level). Consistent with the findings of IH and EH, compared to the male speakers, the female speakers (i.e., reference level) produced AE lower (β = −0.897, SE = 0.193, t = −4.656, $p < 0.001$) and more fronted (β = −0.423, SE = 0.12, t = −3.552, $p < 0.01$). There was a significant interaction between gender and group (GEN1) for the F2 (β = −0.755, SE = −2.07, $p < 0.05$), which indicates that the gender difference was larger for the GEN1 speakers than for the CA speakers. Post-hoc pairwise comparison results confirmed that the female GEN1 speakers, but not the female CA speakers, produced AE significantly more fronted than their male counterparts (β = 0.863, SE = 0.234, t = 3.684, $p < 0.05$). Additionally, a significant gender difference was found in the GEN2 speakers (β = 0.534, SE = 0.113, t = 4.714, $p < 0.01$). The post-hoc test results also showed that, compared to the female CA speakers, female GEN1 speakers' AE was more fronted, although the difference was marginally significant (β = −0.934, SE = 0.283, t = −3.301, $p = 0.056$). No group difference in F2 was observed for the male speakers.

3.3. AE-AEN Split

Figure 5 demonstrates the normalized F1 and F2 values of non-prenasal AE and prenasal AEN across groups. In order to test whether the height (F1) and the frontedness (F2) of these two vowel types were significantly different across groups and genders, we performed linear mixed effects modeling with vowel type (AE, AEN), group (CA, GEN2, GEN1.5, GEN1, KOR), gender (female, male), and the interactions among vowel type, group, and gender as fixed effects and participant and item as random effects. The best fitting models included random intercepts for subject and item with a by-item random slope for gender for the F1 and with a by-subject random slope for vowel for the F2.



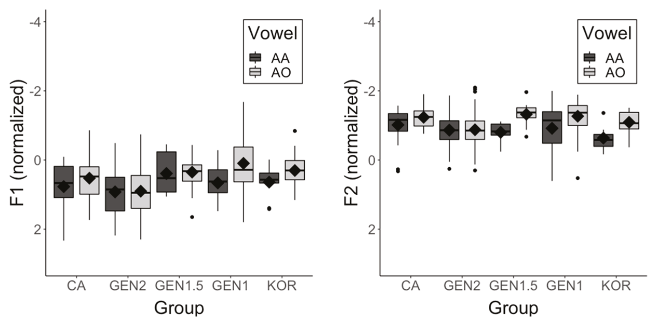
(**a**) (**b**)

Figure 5. AE and AEN across groups: (**a**) Normalized F1; (**b**) Normalized F2.

Results showed that there were main effects of vowel type for both the F1 (β = −0.718, SE = −0.118, t = −6.097, $p < 0.001$) and the F2 (β = 0.515, SE = 0.097, t = 5.315, $p < 0.001$), which suggests that overall AEN was produced significantly higher and more fronted than AE (i.e., reference level). We also found significant interactions between vowel type and all Korean groups in both measures. That is, the difference between AE and AEN was larger for the CA speakers than for the GEN2 (F1: β = 0.794, SE = 0.182, t = 4.356, $p < 0.001$; F2: β = −0.693, SE = 0.172, t = −4.021, $p < 0.001$), the GEN1.5 (F1: β = 0.86, SE = 0.267, t = 3.222, $p < 0.01$; F2: β = −0.531, SE = 0.242, t = −2.19, $p < 0.05$), the GEN1 (F1: β = 1.185, SE = 0.248, t = 4.783, $p < 0.001$; F2: β = −1.001, SE = 0.221, t = −4.529, $p < 0.001$), and the KOR speakers (F1: β = 1.253, SE = 0.369, t = 3.391, $p < 0.001$; F2: β = −1.07, SE = 0.314, t = −3.407, $p < 0.01$). Results of post-hoc pairwise comparisons revealed that, compared to AEN, the CA and the GEN2 speakers produced AE significantly lower (CA: β = 1.537, SE = 0.188, t = 8.179, $p < 0.001$; GEN2: β = 0.743, SE = 0.096, t = 7.712, $p < 0.001$) and more fronted (CA: β = −1.174, SE = 0.169, t = −6.944, $p < 0.001$; GEN2: β = −0.481, SE = 0.086, t = −5.606, $p < 0.001$). The GEN1.5 speaker also showed slightly higher and more fronted AEN than AE, but the difference did not reach statistical significance (F1: β = 0.676, SE = 0.222, t = 3.041, $p = 0.077$; F2: −0.643, SE = −3.333, $p = 0.057$). The late learners (i.e., GEN1 and KOR) did not distinguish the two vowel types. The effect of gender on the F1 and the F2 maintained in the combined AE-AEN data (F1: β = −0.943, SE = 0.178, t = −5.3, $p < 0.001$; F2: β = −0.561, SE = 0.132, t = −4.244, $p < 0.001$). For the F2, there was a marginally significant interaction between gender and group (GEN2) (β = −0.665, SE = 0.324, t = −2.051, $p = 0.05$) and a significant difference between gender and group (GEN1) (β = −1.062, SE = 0.379, t = −2.803, $p < 0.01$). That is, the gender difference was larger for these speakers than for the CA speakers. According to the results of the post-hoc pairwise comparisons, the gender difference was significant for both the GEN2 (β = 0.603, SE = 0.126, t = 4.793, $p < 0.01$) and for the GEN1 speakers (β = 1, SE = 0.235, t = 4.256, $p < 0.01$), while the CA speakers did not show any gender difference. We also found a three-way interaction among vowel type, group (GEN1.5), and gender, suggesting that the interaction between vowel type and group (GEN1.5) showed a different pattern between female and male speakers. We further examined this by running separate models for each gender and found that, while significant interaction between vowel type and group (GEN1.5) appeared in the male data ($\beta = -1.114$, SE = 0.198, t = −5.635, $p < 0.001$), it did not appear in the female data. Given that we only had one female speaker in the GEN1.5 group, these results should be interpreted with caution.

3.4. AA-AO Merger

Figure 6 demonstrates the normalized F1 and F2 values of AA and AO across groups. In order to test whether these two vowels are produced as one category, we performed linear mixed effects modeling with vowel (AA, AO), group (CA, GEN2, GEN1.5, GEN1, KOR), gender (female, male), and the interaction among vowel, group, and gender as fixed effects and participant and item as random effects. The best fitting models included random intercepts for subject and item with a by-subject random slope for vowel for the F1 and with a by-item random slope for gender for the F2.



(**a**) (**b**)

Figure 6. AA and AO across groups: (**a**) Normalized F1; (**b**) Normalized F2.

Results showed that there was a main effect of vowel on both the F1 and the F2, which suggests that overall AA (i.e., reference level) was produced significantly lower and more fronted than AO (F1: $\beta = -0.291$, SE = 0.094, t = −3.102, $p < 0.01$; F2: $\beta = -0.289$, SE = 0.073, t = −3.947, $p < 0.001$). A main effect of group (GEN2) was found ($\beta = 0.221$, SE = 0.094, t = 2.354, $p < 0.05$) for the F2, indicating that overall the GEN2 speakers produced the vowels significantly more fronted than the CA speakers (i.e., reference level). For the F1, there was a marginally significant interaction between vowel and group (GEN1) ($\beta = -0.419$, SE = 0.208, t = −2.014, $p = 0.055$). Post-hoc pairwise comparison results revealed that the GEN1 speakers produced AA significantly lower than AO ($\beta = 0.607$, SE = 0.15, t = 4.051, $p < 0.01$), whereas the CA speakers did not distinguish these vowels. With regard to the F2, we found significant interactions between vowel and group (GEN2) ($\beta = 0.171$, SE = 0.072, t = 2.392, $p < 0.05$) and between vowel and group (GEN1) ($\beta = -0.223$, SE = 0.09, t = −2.468, $p < 0.05$). That is, the difference in F2 between AA and AO of the CA speakers was larger than the GEN2 speakers and smaller than the GEN1 speakers. Post-hoc pairwise comparison results revealed that the CA speakers and the early bilinguals (i.e., GEN2, GEN1.5) produced AA and AO similarly, whereas the late bilinguals distinguished them (GEN1: $\beta = 0.421$, SE = 0.09, t = 4.659, $p < 0.001$; KOR: $\beta = 0.417$, SE = 0.127, t = 3.288, $p < 0.05$). AA and AO did not differ in any measures across groups, except between the GEN2 and the GEN1 speakers in the production of AO (F1: $\beta = 0.707$, SE = 0.155, t = 4.549, $p < 0.01$; F2: $\beta = 0.288$, SE = 0.079, t = 3.629, $p < 0.05$). That is, the GEN2 speakers produced this vowel significantly lower and more fronted than the GEN1 speakers. The GEN2 speakers also produced AO more fronted than the CA speakers, in which the difference approached statistical significance ($\beta = 0.306$, SE = 0.093, t = 3.289, $p = 0.061$). Lastly, the female speakers produced the vowels significantly lower and more fronted than the male speakers (F1: $\beta = -0.291$, SE = 0.094, $t = -3.102$, $p < 0.01$; F2: $\beta = -0.467$, SE = 0.084, $t = -5.55$, $p < 0.001$).

3.5. Fronting of UW, UH, and AH

For the fronting of UW, UH, and AH, we examined the effects of group and gender on the frontedness (F2) of each of these vowels. We performed linear mixed effects modeling with group (CA, GEN2, GEN1.5, GEN1, KOR), gender (female, male), and the interaction between group and gender as fixed effects and participant and item as random effects. In the case of UW, since there was no male token in the data, we did not include the interaction between group and gender as a fixed effect. Additionally, we included the previous segment (coronal, non-coronal) as a fixed effect, since studies have shown that these vowels demonstrate more advanced fronting after coronal consonants than in other phonological environments (D'Onofrio et al. 2019; Podesva et al. 2015). The best-fitting models included random intercepts for subject and item without any random slope.

Results showed that group (GEN2) had an effect on the frontedness of UH ($\beta = 0.581$, SE = 0.171, $t = 3.393$, $p < 0.01$) and AH ($\beta = 0.394$, SE = 0.135, $t = 2.922$, $p < 0.01$). That is, compared to the CA speakers (i.e., reference level), the GEN2 speakers produced these vowels more fronted. No difference was found between the two groups in the production of UW. Lastly, regarding gender, the female speakers produced the three vowel types significantly more fronted than the male speakers (UW: $\beta = -0.431$, SE = 0.199, $t = -2.159$, $p < 0.05$; UH: $\beta = -0.765$, SE = 0.144, $t = -5.318$, $p < 0.001$; AH: $\beta = -0.568$, SE = 0.105, $t = -5.415$, $p < 0.001$). No significant interaction was found between group and gender in UH and AH. Moreover, no effect of previous segment was found in any of the vowels.

4. Discussion

4.1. Effect of Age of Arrival on the Participation in Local Sound Change

In this study we examined the effect of age of arrival on Korean Americans' participation in the California Vowel Shift. We compared the speech of Korean Americans of three generations who clearly differed in the age of arrival to Los Angeles: first-generation (GEN1) (i.e., adulthood), 1.5-generation (GEN1.5) (i.e., late childhood), and second-generation (GEN2) (i.e., early childhood). We predicted that, despite their long residence in the US (average 26.8 years), the GEN1 speakers would show signs of L1 Korean influence in their speech, similar to Korean international students (KOR) who spent less time in the US (average 4 years). On the other hand, younger generation Koreans (i.e., GEN1.5 and GEN2) would perform more similarly to Anglo-Californians (CA) than the KOR speakers and influence from Korean phonology, if any, would appear to a lesser extent for the GEN2 speakers than for the GEN1.5 speakers.

We examined four main patterns of the California Vowel Shift: (1) lowering and retraction of IH, EH, and AE, (2) AE-AEN split, (3) AA-AO merger, and (4) fronting of UW, UH, and AH. For each vowel, we predicted the outcomes based on previous findings of Korean-English late bilinguals and cross-linguistic studies between Korean and English vowels (Baker and Trofimovich 2005; Baker et al. 2002; Trofimovich et al. 2011; Tsukada et al. 2005; Yang 1996; Yoon and Kim 2015).

With regard to IH, as Korean does not have a high front lax vowel, Korean-English late bilinguals tend to merge this vowel to IY which is almost identical to the Korean /i/ (Baker and Trofimovich 2005; Yang 1996). Thus, we predicted that, if influence from Korean phonology occurs, Korean Americans would merge IH with IY and, thus, would not participate in the lowering and retraction of IH. Results showed that the early bilinguals (i.e., GEN2 and GEN1.5) aligned with the CA speakers in that they distinguished the IY-IH contrast using both vowel height and frontedness. On the other hand, the late bilinguals (i.e., GEN1 and KOR) patterned similarly to each other; they only used vowel frontedness to distinguish the contrast. Although the GEN1 and the KOR speakers did not completely merge IH with IY, which would have been a strong indication of Korean influence, their IH approached IY in the vowel space (see Figure 3). Indeed, when comparing the difference of the average normalized

F2 (i.e., vowel frontedness) between IY and IH across groups, the late bilinguals demonstrated smaller differences (GEN1: 0.11, KOR: 0.44) than the CA speakers (1.08) and the early bilinguals (GEN2: 0.89, GEN1.5: 1.35). Moreover, we found that both the GEN1 and the KOR speakers produced IH more fronted than the CA speakers. The GEN2 speakers also produced this vowel more fronted than the CA speakers. However, unlike the GEN1 and the KOR speakers, we do not believe that this is due to influence from Korean phonology, given that the GEN2 speakers additionally produced this vowel lower than the CA speakers. If the Korean /i/ had an effect on GEN2 speakers' IH, it would have demonstrated higher vowel height than CA speakers' IH. In fact, among the five groups, GEN2 speakers' IH was produced with the lowest vowel height (see Footnote 12). Rather than influence from Korean phonology, GEN2 speakers' divergence from the CA speakers could be explained through the nature of their vowel space. The vowel space of the GEN2 speakers was overall more fronted than that of the CA speakers.

Regarding EH and AE, we made different predictions for the two vowels. While Korean does not have any vowel that acoustically overlaps with either EH or AE, Korean-English bilinguals often identify both vowels as the Korean /e/ (Baker et al. 2002; Trofimovich et al. 2011), which is positioned higher in the vowel space (Baker and Trofimovich 2005; Yoon and Kim 2015). However, instead of assimilating the merged EH-AE category to the Korean /e/, Korean-English bilinguals, especially late bilinguals, tend to demonstrate two categories in the front mid/low region of the vowel space: the Korean /e/ and a single EH-AH category in which AE merges with EH (Baker and Trofimovich 2005). Therefore, we predicted that, if influence from Korean occurs, Korean Americans would merge AE with EH; their EH may demonstrate lowering and retraction, but their AE would not, because it would be positioned higher in the vowel space due to the merger with EH. Contrary to our prediction, the Korean Americans in our study maintained the EH-AE contrast regardless of their age of arrival to the US. As in the case of IH, we found a different pattern between the early bilinguals and the late bilinguals. The GEN2 and the GEN1.5 speakers performed like the CA speakers in that they produced EH and AE distinctly using both vowel height and frontedness. On the other hand, the GEN1 speakers and the KOR speakers kept the EH-AE contrast using only one measure (i.e., vowel height).

While the GEN1 speakers distinguished the EH-AH contrast, they produced both vowels higher and more fronted than the CA speakers, indicating that these speakers did not participate in the lowering and retraction of EH and AE. The GEN2 speakers also produced EH more fronted than the CA speakers, but no difference in vowel height was found between the two groups. As mentioned above, we believe that is due to their overall more fronted vowel space compared to the CA speakers. Although participants' birth year was not the main focus of our study, it is worth pointing out that the GEN1 speakers were overall older than the other groups (see Table 1). Although we formulated our predictions with the assumption that the merger between the Korean /e/ and /ɛ/ is established across most age groups (Kang 2014; Jang et al. 2015), it is possible that the GEN1 speakers arrived in the US during the time when the merger was still in progress. For instance, Yang (1996) demonstrated that Korean male adults in the 1990s maintained the distinction between the Korean /e/ and /ɛ/, while Korean female adults produced them indistinguishably, suggesting that the Korean /e/-/ɛ/ merger was still in progress during this period and female speakers led the change. Most of the GEN1 speakers in our study immigrated to the US during the 1970s and the 1980s. After a long period of time away from Korea, it is likely that the GEN1 speakers do not participate in sound changes in Korea that are still in progress or that were established after they left. In fact, linguistic conservatism has often been observed in diasporic communities (Johannessen and Laake 2015; Parodi 2014; Polinsky 2018). Thus, it is possible that the GEN1 speakers kept the Korean /e/-/ɛ/ contrast that they brought with them, which may have affected their production of (unmerged) EH and AE.

With regard to the KOR speakers, given that these speakers were much younger than the GEN1 speakers and left Korea recently (see Table 1), it is unlikely that they maintain the Korean /e/-/ɛ/ contrast. When comparing the difference of the average normalized F1 (i.e., vowel height) between EH and AE across groups, the KOR speakers demonstrated a smaller difference (0.42), compared to the CA speakers (0.75) and the early bilinguals (GEN2: 0.64, GEN1.5: 0.66). On the other hand, GEN1 speakers' vowel height difference between EH and AE (0.71) was comparable to that of the CA speakers and the early bilinguals. Thus, it appears that both the GEN1 and the KOR speakers demonstrate influence from Korean phonology when producing EH and AE, but in a different way. While the GEN1 speakers assimilate EH and AE to the Korean /e/ and /ɛ/, respectively, the KOR speakers acquire the EH-AE contrast using vowel height, but do so less consistently than the CA speakers, similar to the case of the IY-IH contrast. To confirm this, future research should examine GEN1 speakers' and KOR speakers' realization of both English and Korean vowels.

In this study, we examined whether Korean Americans produce AE differently based on the nasality of the following consonant (i.e., AE-AEN split). Korean does not demonstrate a systematic split between non-prenasal and prenasal vowels, thus, we predicted that, if influence from Korean phonology occurs, Korean Americans would not distinguish AE and AEN. Our results showed that late bilinguals (i.e., GEN1 and KOR) and the GEN1.5 speakers did not distinguish AE and AEN. The GEN2 speakers, on the other hand, produced AE and AEN distinctly using both vowel height and frontedness, but the difference between these two vowels were smaller than for the CA speakers. This finding suggests that the GEN2 speakers participate in the AE-AEN split, but to a lesser extent than the CA speakers.

Regarding the low back vowels AA and AO, Korean does not have a vowel that acoustically overlaps with any of these vowels. The closest vowels in Korean would be /a/ (low central) and /ʌ/ (near-low back) (Baker et al. 2002; Trofimovich et al. 2011; Tsukada et al. 2005). Thus, we predicted that if influence from Korean occurs, they would either assimilate both vowels to the Korean /a/ (Outcome 1: Participation in AA-AO merger but more fronted than expected) or distinguish them by assimilating AA to the Korean /a/ and assimilating AO to the Korean /ʌ/ (Outcome 2: No participation in AA-AO merger). Results showed that the late bilinguals (i.e., GEN1 and the KOR) produced AA more fronted than AO. Additionally, the GEN1 speakers produced AA lower than AO. These findings suggest that the late bilinguals did not participate in the AA-AO merger, most likely because they assimilated AA to the Korean /a/ and assimilated AO to the Korean /ʌ/ (i.e., Outcome 2). The finding that these speakers produced AO indistinctly from AH, which is the closest vowel to the Korean /ʌ/ (see Section 3.1), supports the possibility that they assimilated AO to the Korean /ʌ/. As for the GEN2 and the GEN1.5 speakers, they patterned like the CA speakers in that they showed an overlap between AA and AO. In the case of the GEN2 speakers, the AA-AO merger was even stronger than the CA speakers, suggesting that the GEN2 speakers may be in a more advanced stage of the AA-AO merger. However, compared to the CA speakers their productions were overall more fronted. Although the results seem to be pointing toward Outcome 1 (i.e., assimilation to the Korean /a/), this is unlikely. Compared to the CA speakers, GEN2 speakers' AO was more fronted, but not lower, which would have been the case if the GEN2 speakers assimilated the merged AA-AO category to the Korean /a/. As shown in Figure 3, AA is positioned in the low-back area in GEN2 speakers' vowel space, whereas for the late bilinguals it is positioned in the low-mid area between AE and AO. If the GEN2 speakers assimilated AA to the Korean /a/, which is what we believe happened in the speech of the late bilinguals, they would have shown similar patterns as the late bilinguals. Thus, as in the case of the front vowels, GEN2 speakers' overall fronted vowel space seems to be a more plausible explanation to their divergence from the CA speakers.

Lastly, with regard to the high back vowels UW and UH, Korean-English bilinguals tend to assimilate both vowels to the Korean /u/ which is more back (Baker and Trofimovich 2005). Thus, we predicted that, if influence from Korean phonology occurs, Korean Americans would merge UW and UH and produce them more back than the CA speakers. Similarly, AH would be produced more back than the CA speakers due to influence from the Korean /ʌ/. Our data showed that the late bilinguals produced UW and UH indistinguishably, suggesting a strong influence from Korean phonology. However, unlike what we expected, they did not produce the merged UW-UH more back than CA speakers. One possible explanation is that, rather than to the Korean /u/, these speakers may have assimilated the UW-UH category to the Korean /ɨ/ which is acoustically more similar to these vowels. Future research examining late bilinguals' combined L1 and L2 vowel space would help confirm this. Unlike the late bilinguals, the early bilinguals aligned with the CA speakers in that they maintained the UW-UH contrast using vowel height[12]. Compared to the CA speakers, only the GEN2 speakers demonstrated more fronted UH and AH. As mentioned above, we believe that this is due to their more fronted vowel space.

Overall, our data showed a clear distinction between the GEN2 and the GEN1.5 speakers (i.e., early bilinguals), on the one hand, and the GEN1 speakers (i.e., late bilinguals), on the other, confirming an effect of age of arrival to the US on Korean Americans' realization of English vowels. Similar to the KOR speakers, the GEN1 speakers did not distinguish the front vowels contrasts IY-IH and EH-AE, using the same strategies as the CA speakers, and failed to maintain the back vowel contrasts UW-UH and AH-AO. Moreover, they did not participate in the California Vowel Shift, which is mostly likely due to influence from their L1 Korean. Unlike the late bilinguals, the early bilinguals successfully maintained the four vowel contrasts using the same phonetic strategies as the CA speakers. Moreover, the CA speakers and the early bilinguals demonstrated horizontally narrower and vertically more expanded vowel space than the late bilinguals (see Figure 3). This indicates that these speakers followed the linguistic trend of California English which is characterized by a horizontal compression of vowel space (D'Onofrio et al. 2019). However, the early bilinguals did not demonstrate a complete convergence toward the CA speakers, especially when producing non-prenasal AE and prenasal AEN. While the GEN2 speakers distinguished the two vowel types, their split was less pronounced than the CA speakers. The GEN1.5 speakers, on the other hand, did not demonstrate a systematic distinction between AE and AEN, following the patterns of the late bilinguals. Less pronounced or lack of AE-AEN split among Korean Americans has also been reported in other studies (Cheng 2016; Lee 2016). Since Korean does not have AE-AEN split, this finding suggests that Korean phonology has an effect on early bilinguals' production of AE and AEN and that the GEN1.5 speakers demonstrate a stronger influence from Korean phonology than the GEN2 speakers due to their later exposure to California English (i.e., age effect).

4.2. Second-Generation Korean Americans' Divergent Participation in the California Vowel Shift

In the case of the GEN2 speakers, apart from the less pronounced AE-AEN split, we found that these speakers additionally demonstrated an overall more fronted realization of the vowels than the CA speakers. Except for front vowel retraction, all the patterns of the California Vowel Shift examined in this study were observed in GEN2 speakers' speech (i.e., front vowel lowering, back vowel fronting, AE-AEN split, AA-AO merger). In fact, in certain aspects, the GEN2 speakers seemed to be in a more advanced stage of the California Vowel Shift than the CA speakers (i.e., IH-lowering, AA-AO merger, UH- and AH-fronting). These findings, along with the less pronounced AE-AEN split, are highly consistent with those of Korean Americans in Berkeley (Cheng 2016), which suggests that Korean Americans in Southern and Northern California may share similar patterns. Based on visual inspection of participants' vowel space in Figure 3, the GEN2 speakers seemed to demonstrate the narrowest vowel space across groups. Thus, it is possible that GEN2 speakers' fronted vowel space occurs in combination with more advanced horizontal compression than the CA speakers.

---

[12] It is noteworthy that the CA speakers and the early bilinguals did not use vowel frontedness to distinguish the UW-UH contrast, whereas they used both vowel height and frontedness when distinguishing other contrasts (see Table 2). We suspect that this is linked to the lack of a main effect of previous consonant on the production of UW and UH. The findings of this study differed from previous research (D'Onofrio et al. 2019; Hall-Lew 2009; Podesva et al. 2015) in that the production of UW and UH was not conditioned by the phonological context that encourages fronting (i.e., post-coronal position). Although our data do not have enough tokens in post-coronal position to further examine its effect on individual speakers, these findings seem to indicate that the fronting of UW and UH is well established among the CA speakers and the early bilinguals. A similar claim has been made by Hall-Lew (2009, 2011) that back vowel fronting is nearing completion in Northern California.

While further examination of GEN2 speakers' holistic vowel space (e.g., area and dispersion) should be carried out, it appears that there is a link between GEN2 speakers' narrow vowel space and their pronounced back vowel fronting. Pronounced back vowel fronting has also been found in other Asian American groups. For instance, Hall-Lew (2009, 2011) demonstrated that Chinese Americans in San Francisco may be in a more advanced stage of back vowel fronting than Anglo-Californians. Similarly, Cheng (2016) found that, apart from the Korean Americans, South Asians also demonstrated more pronounced UH-fronting than Anglo-Californians. Thus, it is possible that some Asian Americans in California collectively demonstrate stronger participation in back vowel fronting than Anglo-Californians to express their pan-ethnic Asian American identity. According to Wei (1993, p. 1), being Asian American "implies that there can be a communal consciousness and a unique culture that is neither Asian or American, but Asian American." US-born Asian Americans often experience microaggressions challenging their American-ness due to their phenotypic traits that are distinct from the mainstream Americans (i.e., Anglo-Americans) (Lee 2019). The shared racialization experiences, which contradicts the covert oppression exerted upon Asian Americans behind the model minority stereotype (e.g., docile, hard-working, good citizens) (Chou and Feagin 2010; Kawai 2005; Lee 2019), may lead some Asian Americans to overemphasize their American-ness using linguistic resources. In other words, the pronounced back vowel fronting observed in the GEN2 speakers may be a result of the speakers overcompensating for their perceived un-American-ness by taking the back vowel fronting of the California Vowel Shift even further than the CA speakers. This may eventually cause for the front vowels to be pushed forward in order to maintain sufficient perceptual contrasts between front and back vowels (Lindblom 1990; Lindblom and Engstrand 1989). Future research should examine the social meanings of back vowel fronting and the relationship between the degree of back vowel fronting and pan-ethnic Asian American membership across different Asian American groups, as well as its effect on the realization of front vowels.

It is important to note that, unlike the GEN2 speakers, the GEN1.5 speakers did not demonstrate pronounced back vowel fronting or an overall fronted vowel space. If we extend our argument from above, it is possible that the GEN1.5 speakers may not feel the need to overemphasize their American-ness through back vowel fronting in the same way as the GEN2 speakers, since GEN1.5 speakers often demonstrate a strong affiliation to Korean cultures as part of their dual identity (Kim and Stodolska 2013). Thus, it is likely that GEN1.5 speakers identify themselves more strongly as Koreans or Korean Americans than Asian Americans. Due to the small sample size (N = 4), it is premature to make an assumption on GEN1.5 speakers' speech behaviors. Future research should include a balanced number of GEN1.5 and GEN2 speakers to test whether their English vowels systematically differ from each other and whether their pan-ethnic Asian American identity in relation to their Korean or Korean American identity has an effect on their realization of English vowels.

Another possible explanation to GEN2 speakers' fronted vowel space is the social meaning associated with gender in Korean culture. Cross-linguistically, female speakers have higher fundamental frequency (F0) and formant frequencies than male speakers due to differences in their vocal anatomy (Escudero et al. 2009; Jacewicz et al. 2007; Pisanski et al. 2016; Simpson 2002; Yoon and Kim 2015). Thus, compared to male speakers, female speakers generally have a higher-pitched voice and produce vowels with lower height (i.e., higher F1) and more fronted (i.e., higher F2). In this study, we normalized participants' formant frequencies in order to examine gender effects on English vowel production while controlling for physiological differences between female and male speakers.

As demonstrated in Figure 3, the vowel space of CA female and male speakers largely overlapped in the front-back dimension[13], whereas clear gender differences were observed across Korean groups,

[13] The only gender difference in the CA group was found in the vowel height of EH. That is, the CA female speakers produced EH lower than the male speakers. Kennedy and Grama (2012) found similar results in that female and male Californians in Santa Barbara (Southern California) differed in the height of AE (i.e., women produced it lower than men), but did not show any significant difference in vowel frontedness. Since women in general are leaders of sound changes (Coates 1993; Milroy

especially among the GEN2 and the GEN1 speakers. That is, even after normalizing formant frequencies, the Korean female speakers produced English vowels more fronted than the Korean male speakers. These findings suggest that Korean female speakers shift their vowel space forward as a way to express their femininity. Femininity is indexed differently across cultures. For instance, in American culture, women use creaky voice to enhance their female desirability (Pennock-Speck 2005; Yuasa 2010), whereas Japanese women use high-pitched voice to sound cute, young, and charming (Van Bezooijen 1995; Ohara 1998; Yuasa 2010). Although to a lesser degree than in Japanese culture, Korean women also use high-pitched voice to express femininity (Ohara 1998; Puzar and Hong 2018). High-pitched voice is a characteristic of performed winsomeness called aegyo, which is the cutified and infantilized figuration of femininity in Korean culture[14] (Puzar and Hong 2018). Due to the close relationship between F0 and formants, it is likely that performers of aegyo also produce fronted vowel space. According to Pisanski et al. (2016), speakers across genders and cultures modulate their vocal tract length and F0 to imitate a physically large and small body size. That is, they shorten their vocal tract and increase their F0 to sound physically small and do the opposite to sound physically large. Thus, it is possible that the GEN2 female speakers move their vowel space forward to express Korean femininity. Here we would like to emphasize that among the four Korean groups (i.e., GEN2, GEN1.5, GEN1, and KOR), gender differences surfaced most systematically in the vowels of the GEN2 and the GEN1 speakers and that the front region of GEN2 speakers' vowel space largely overlapped with that of the GEN1 speakers. That is, GEN2 speakers' fronted vowel space reflects features that are present in both the CA speakers (i.e., horizontal compression and vertical expansion of the vowel space) and the Korean speakers (i.e., more fronted vowel space among female than male speakers), particularly those of their parents' generation. Although evidence of individual vowels disfavors the possibility of influence from Korean phonology, the findings suggest that the vowel space of the female GEN2 speakers is moving forward to align with the front region of the female GEN1 speakers. Studies have shown that children of immigrants who acquire the majority language natively may use ethnolectal features as additional linguistic resources to mark social meanings (e.g., association with ethnicity) (Cheshire et al. 2011; Clyne et al. 2001; Gnevsheva 2020). While ethnolectal features may originate from first-generation immigrants' foreign-accented speech, in second generation they may be reallocated for sociolinguistic purposes (Clyne et al. 2001; Gnevsheva 2020; Hoffman and Walker 2010). Thus, it is possible that the female GEN2 speakers shift their entire vowel space forward to index their intersecting ethnic and gender identities (i.e., cute and charming Korean female persona). The vowel space shift may occur independently or in combination with pronounced back vowel fronting to additionally express their pan-ethnic Asian American identity, as proposed above. Future research should examine intra-speaker variation of GEN2 speakers' vowel productions (e.g., style-shifting) to understand the social meanings of their fronted vowel space. Moreover, a perceptual study should be accompanied to examine whether such social meanings are shared by the Korean American community.

5. Conclusions

In this study, we examined Korean Americans' participation in the California Vowel Shift. Although the first-generation Korean Americans had spent a much longer time in the US than the Korean international students, influence from Korean still persisted in their speech. On the other hand, Korean Americans who were born and raised in Los Angeles (i.e., second-generation) or those who

---

and Milroy 1985; Labov 1990; Trudgill 1972), it appears that the lowering of mid and low front vowels EH and AE is still in progress in California English, whereas changes in the front-back dimension (i.e., retraction of front vowels and fronting of back vowels) may be nearing stability for the CA speakers.

[14] According to Puzar and Hong (2018), aegyo is not a direct emulation of child behaviors, but a performative repertoire of secondary infantilisation (Goffman 1979, pp. 72–77) used for various purposes (e.g., playfulness, seduction, negotiation, pleasing superiors). Thus, performers of aegyo, particularly young women, use this speech style to "negotiate the imbalance of power within patriarchal, androcentric and ageist/gerontocratic environments" (Puzar and Hong 2018). Similar concepts exist in other East Asian cultures, such as sajiao in China (Farris 1994) and kawaii in Japan (Brown 2011; Madge 1998).

came to the US during childhood (i.e., 1.5-generation) demonstrated most patterns of the California Vowel Shift. However, divergence from the Anglo-Californians was observed in their production of prenasal and non-prenasal /æ/. The 1.5-generation speakers did not systematically distinguish the two vowel types, similar to the late bilinguals. The second-generation speakers demonstrated a split-/æ/ system, but it was less pronounced than for the Anglo-Californians. These findings suggests that age of arrival has a strong effect on immigrant minority speakers' participation in local sound change.

Our findings also showed that the second-generation Korean Americans, in particular the female speakers, demonstrated an overall more fronted realization of the vowels than the Anglo-Californians. Second-generation Korean Americans' fronted vowel space reflected features that were present in both the Anglo-Californians (i.e., horizontal compression and vertical expansion of the vowel space) and the Korean speakers, particularly those of the first-generation Korean Americans (i.e., more fronted vowel space among female than male speakers). These findings suggest that second-generation Korean Americans may shift their vowel space forward to express their intersecting gender, racial, and ethnic identities. Future research should examine the social meanings of the fronting of vowel space.

Author Contributions: Conceptualization, J.Y.K.; Data curation, N.W.; Formal analysis, J.Y.K.; Investigation, J.Y.K. and N.W.; Methodology, J.Y.K. and N.W.; Visualization, J.Y.K. and N.W.; Writing—original draft, J.Y.K.; Writing—review & editing, J.Y.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

Ahn, Sang-Cheol, and Gregory K. Iverson. 2007. Structured imbalances in the emergence of the Korean vowel system. In *Historical Linguistics 2005: Selected Papers from the 17th International Conference on Historical Linguistics*. Edited by Joseph C. Salmons and Shannon Dubenion-Smith. Amsterdam and Philadelphia: John Benjamins, pp. 275–93.

Baker, Wendy, and Pavel Trofimovich. 2005. Interaction of Native- and Second-Language Vowels System(s) in Early and Late Bilinguals. *Language and Speech* 48: 1–27. [CrossRef]

Baker, Wendy, Pavel Trofimovich, Molly Mack, and James E. Flege. 2002. The Effect of Perceived Phonetic Similarity on Non-Native Sound Learning by Children and Adults. In *Proceedings of the 26th Annual Boston University Conference on Language Development*. Edited by Barbora Skarabela, Sarah Fish and Anna H.-J. Do. Somerville: Cascadilla Press, pp. 36–47.

Bates, Douglas, Martin Maechler, Bon Bolker, and Steve Walker. 2015. Fitting Linear Mixed-effects Models Using lme4. *Journal of Statistical Software* 67: 1–48. [CrossRef]

Best, Catherine T., and Michael D. Tyler. 2007. Nonnative and Second-language Speech Perception: Commonalities and Complementarities. In *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*. Edited by Ocke-Schwen Bohn and Murray J. Munro. Amsterdam and Philadelphia: John Benjamins, pp. 13–34.

Boberg, Charles. 2004. Ethnic Patterns in the Phonetics of Montreal English. *Journal of Sociolinguistics* 8: 538–68. [CrossRef]

Boersma, Paul, and David Weenink. 2020. Praat: Doing Phonetics by Computer. Computer Program. Version 5.1.31. Available online: http://www.praat.org/ (accessed on 3 July 2020).

Brown, Janice. 2011. Re-framing "Kawaii": Interrogating Global Anxieties Surrounding the Culture of 'Cute' in Japanese Art and Consumer Products. *International Journal of the Image* 1: 1–10. [CrossRef]

Carter, Phillip E., Lydda L. Valdez, and Nandi Sims. 2020. New Dialect Formation through Language Contact: Vocalic and Prosodic Developments in Miami English. *American Speech* 95: 119–48. [CrossRef]

Casillas, Joseph V., and Miquel Simonet. 2016. Production and Perception of the English/æ/-/ɑ/Contrast in Switched-dominance Speakers. *Second Language Research* 32: 171–95. [CrossRef]

Chang, Charles B., Yao Yao, Erin F. Haynes, and Russell Rhodes. 2011. Production of Phonetic and Phonological Contrast by Heritage Speakers of Mandarin. *The Journal of the Acoustical Society of America* 129: 3964–80. [CrossRef] [PubMed]

Cheng, Andrew. 2016. A Survey of English Vowel Spaces of Asian American Californians. *UC Berkeley Phonetics and Phonology Lab Annual Report* 12: 348–84.

Cheshire, Jenny, Paul Kerswill, Sue Fox, and Eivind Torgersen. 2011. Contact, the Feature Pool and the Speech Community: The Emergence of Multicultural London English. *Journal of Sociolinguistics* 15: 151–96. [CrossRef]

Chou, Rosalind S., and Joe R. Feagin. 2010. *The Myth of the Model Minority: Asian Americans Facing Racism*. Boulder: Paradigm Publishers.

Clarke, Sandra, Ford Elms, and Amani Youssef. 1995. The Third Dialect of English: Some Canadian Evidence. *Language Variation and Change* 7: 209–28. [CrossRef]

Clyne, Michael G., Edina Eisikovits, and Laura F. Tollfree. 2001. Ethnic varieties of Australian English. In *English in Australia*. Edited by David Blair and Peter Collins. Amsterdam: John Benjamins, pp. 223–38.

Coates, Jennifer. 1993. *Women, Men, and Language: A Sociolinguistic Account of Gender Differences in Language*, 2nd ed. London: Longman.

D'Onofrio, Annette, Penelope Eckert, Robert J. Podesva, Teresa Pratt, and Janneke Van Hofwegen. 2016. The Low Vowels in California's Central Valley. *Publication of the American Dialect Society* 101: 11–32. [CrossRef]

D'Onofrio, Annette, Teresa Pratt, and Janneke Van Hofwegen. 2019. Compression in the California Vowel Shift: Tracking Generational Sound Change in California's Central Valley. *Language Variation and Change* 31: 193–217. [CrossRef]

Eckert, Penelope. 1989. *Jocks and Burnouts: Social Categories and Identity in the High School*. New York: Teachers College Press.

Eckert, Penelope. 2008. Where Do Ethnolects Stop? *International Journal of Bilingualism* 12: 25–42. [CrossRef]

Escudero, Paola, Paul Boersma, Adréia Schurt Rauber, and Ricardo A. H. Bion. 2009. A Cross-dialect Acoustic Description of vowels: Brazilian and European Portuguese. *The Journal of the Acoustical Society of America* 126: 1379–93. [CrossRef]

Farris, Catherine S. 1994. A Semiotic Analysis of Sajiao as a Gender Marked Communication Style in Chinese. In *Unbound Taiwan: Closeups from a Distance, Select Papers 8*. Edited by Marshall Johnson and Fred Y. L. Chiu. Chicago: Center for East Asian Studies, University of Chicago, pp. 2–29.

Flege, James. E. 1995. Second-language Speech Learning: Theory, Findings, and Problems. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Edited by Winifred Strange. Timonium: York Press, pp. 229–73.

Flege, James. E. 1999. Age of Learning and Second-language Speech. In *Second Language Acquisition and the Critical Period Hypothesis*. Edited by David Birdsong. Mahwah: Lawrence Erlbaum Press, pp. 101–32.

Flege, James E., and Ian R. A. MacKay. 2011. What Accounts for "Age" Effects on Overall Degree of Foreign Accent? In *Achievements and Perspectives in the Acquisition of Second Language Speech. New Sounds*. Edited by Magdalena Wrembel, Malgorzata Kul and Katarzyna Dziubalska-Kołaczyk. Bern: Peter Lang, vol. 2, pp. 65–82.

Flege, James. E., Murray Munro, and Ian R. A. MacKay. 1995. Factors Affecting Strength of Perceived Foreign Accent in a Second Language. *Journal of the Acoustical Society of America* 97: 3125–34. [CrossRef]

Flege, James. E., Ocke-Schewen Bohn, and Sunyoung Jang. 1997. Effects of Experience on Non-native Speakers' Production and Perception of English Vowels. *Journal of Phonetics* 25: 437–70. [CrossRef]

Flemming, Edward. 1996. Evidence for Constraints on Contrast: The Dispersion Theory of Contrast. *UCLA Working Papers in Phonology* 1: 86–106.

Fought, Carmen. 1999. A Majority Sound Change in a Minority Community:/u/-Fronting in Chicano English. *Journal of Sociolinguistics* 3: 5–23. [CrossRef]

Fought, Carmen. 2003. *Chicano English in Context*. New York: Macmillan.

Gnevsheva, Ksenia. 2020. The Role of Style in the Ethnolect: Style-Shifting in the Use of Ethnolectal Features in the First- and Second-generation Speakers. *International Journal of Bilingualism*, 1–20. [CrossRef]

Goffman, Erving. 1979. *Gender Advertisements*, 2nd ed. Cambridge: Harvard University Press.

Gordon, Matthew. 2011. Methodological and Theoretical Issues in the Study of Chain Shifting. *Language and Linguistics Compass* 5: 784–94. [CrossRef]

Hagiwara, Robert. 1997. Dialect Variation and Formant Frequency: The American English Vowels Revisited. *Journal of the Acoustical Society of America* 102: 655–8. [CrossRef]

Hall-Lew, Lauren. 2009. Ethnicity and Sound Change in San Francisco English. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 35: 111–22. [CrossRef]

Hall-Lew, Lauren. 2011. The Completion of a Sound Change in California English. *Proceedings of the International Congress of Phonetic Sciences* 17: 807–10.

Hall-Lew, Lauren, and Rebecca L. Starr. 2010. Beyond the 2nd Generation: English Use Among Chinese Americans in the San Francisco Bay Area. *English Today* 26: 12–19. [CrossRef]

Hall-Lew, Lauren, Amanda Cardoso, Yova Kemenchedjieva, Kieran Wilson, Ruaridh Purse, and Julie Saigusa. 2015. San Francisco English and the California Vowel Shift. Paper presented at 18th International Conference of the Phonetic Sciences, Glasgow, UK, 10–14 August 2015.

Hinton, Leanne, Birch Moonwomon, Sue Bremner, Herb Luthin, Mary Van Clay, Jean Lerner, and Hazel Corcoran. 1987. It's Not Just the Valley Girls: A study of California English. Paper presented at Thirteenth Annual Meeting of the Berkeley Linguistics Society, Berkeley, CA, USA, 14–16 February 1987; pp. 117–28.

Hoffman, Michol F. 2010. The Role of Social Factors in the Canadian Vowel Shift: Evidence from Toronto. *American Speech* 85: 121–40. [CrossRef]

Hoffman, Michol F., and James A. Walker. 2010. Ethnolects and the City: Ethnic Orientation and Linguistic Variation in Toronto English. *Language Variation and Change* 22: 37–67. [CrossRef]

Ito, Rika. 2010. Accommodation to the Local Majority Norm by Hmong Americans in the Twin Cities, Minnesota. *American Speech* 85: 141–62. [CrossRef]

Jacewicz, Ewa, Robert Allen Fox, and Joseph Salmons. 2007. Vowel Space Areas across Dialects and Gender. In *Proceedings of the 16th International Congress of Phonetic Sciences*. Edited by Jürgen Trouvain and William John Barry. Saarbrücken: University of Saarland, pp. 1465–68.

Jang, Hyejin, Jiyoung Shin, and Hosung Nam. 2015. Aspects of Vowels by Ages in Seoul Dialect. *Studies in Phonetics, Phonology, and Morphology* 21: 341–58. [CrossRef]

Jia, Gisela, and Doris Aaronson. 2003. A Longitudinal Study of Chinese Children and Adolescents Learning English in the United States. *Applied Psycholinguistics* 24: 131–61. [CrossRef]

Johannessen, Janne Bondi, and Signe Laake. 2015. On two myths of the Norwegian language in America: Is it old-fashioned? Is it approaching the written Bokmål standard? In *Germanic Heritage Languages in North America*. Edited by Janne Bondi Johannessen and Joseph C. Salmons. Amsterdam and Philadelphia: John Benjamins, pp. 299–322.

Kang, Yoonjung. 2014. A Corpus-based Study of Positional Variation in Seoul Korean Vowels. *Japanese/Korean Linguistics* 23: 1–20.

Kang, Jieun, and Eun Jong Kong. 2016. Static and Dynamic Spectral Properties of the Monophthong Vowels in Seoul Korean: Implication on Sound Change. *Phonetics and Speech Sciences* 8: 39–47. [CrossRef]

Kawai, Yuko. 2005. Stereotyping Asian Americans: The Dialectic of the Model Minority and the Yellow Peril. *The Howard Journal of Communications* 16: 109–30. [CrossRef]

Kennedy, Robert, and James Grama. 2012. Chain Shifting and Centralization in California Vowels: An Acoustic Analysis. *American Speech* 87: 39–56. [CrossRef]

Kim, Jungeun, and Monika Stodolska. 2013. Impacts of Diaspora Travel on Ethnic Identity Development among 1.5 Generation Korean American College Students. *Journal of Tourism and Cultural Change* 11: 187–207. [CrossRef]

Kuhl, Patricia K., Karern A. Williams, Francisco Lacerda, Kenneth N. Stevens, and Björn Lindblom. 1992. Linguistic Experience Alters Phonetic Perception in Infants by 6 Months of Age. *Science* 255: 606–8. [CrossRef]

Kuznetsova Alexandra, Brockhoff Per B., and Christensen Rune H. B. 2017. LmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82: 1–26. [CrossRef]

Kwak, Chung-gu. 2003. The Vowel System of Contemporary Korean and Direction of Change. *Journal of Korean Linguistics* 41: 59–91.

Labov, William. 1972. *Language in the Inner City: Studies in Black English Vernacular*. Philadelphia: University of Pennsylvania Press.

Labov, William. 1990. The Intersection of Sex and Social Class in the Course of Linguistic Change. *Language Variation and Change* 2: 205–54. [CrossRef]

Labov, William. 2001. *Principles of Linguistic Change: Social Factors*. Malden: Wiley-Blackwell, vol. 2.

Labov, William. 2010. *Principles of Linguistic Change: Cognitive and Cultural Factors*. Malden: Wiley-Blackwell, vol. 3.

Labov, William, Sharon Ash, and Charles Boberg. 2006. *The Atlas of North American English: Phonetics, Phonology, and Sound Change*. New York: Mouton de Gruyter.

Lee, Hikyong. 2000. Korean Americans as Speakers of English: The Acquisition of General and Regional Features. Ph.D. dissertation, University of Pennsylvania, Philadelphia, PA, USA.

Lee, Jinsok. 2016. The Participation of a Northern New Jersey Korean American Community in Local and National Language Variation. *American Speech* 91: 327–60. [CrossRef]

Lee, Jess. 2019. Many Dimensions of Asian American Pan-ethnicity. *Sociology Compass* 13: 1–16. [CrossRef]

Lee, Ki-Moon, and S. Robert Ramsey. 2011. *A History of the Korean Language*. Cambridge: Cambridge University Press.

Lee, Hyangwon, Woobong Shin, and Jiyoung Shin. 2017. A Sociophonetic Study on High/Mid Back Vowels in Korean. *Phonetics and Speech Sciences* 9: 39–51.

Lenth, Russell. 2020. Emmeans: Estimated Marginal Means, aka Least-Squares Means. R Package Version 1.5.0. Available online: https://CRAN.R-project.org/package=emmeans (accessed on 8 September 2020).

Liljencrants, Johan, and Björn Lindblom. 1972. Numerical Simulation of Vowel Quality Systems: The Role of Perceptual Contrast. *Language* 48: 839–62. [CrossRef]

Lindblom, Björn. 1990. Explaining Phonetic Variation: A Sketch of the H&H Theory. In *Speech Production and Speech Modelling*. Edited by William J. Hardcastle and Alain Marchal. Dordrecht: Springer, pp. 403–39.

Lindblom, Björn, and Olle Engstrand. 1989. In What Sense is Speech Quantal? *Journal of Phonetics* 17: 107–21. [CrossRef]

Lloyd-Smith, Anika, Marieke Einfeldt, and Tanja Kupisch. 2020. Italian-German Bilinguals: The Effects of Heritage Language Use on Accent in Early-Acquired Languages. *International Journal of Bilingualism* 24: 289–304. [CrossRef]

Lobanov, Boris. 1971. Classification of Russian Vowels Spoken by Different Speakers. *Journal of the Acoustical Society of America* 49: 606–8. [CrossRef]

Madge, Leila. 1998. Capitalizing on 'Cuteness': The Aesthetics of Social Relations in a New Postwar Japanese Order. *Japanstudien* 9: 155–74. [CrossRef]

Martinet, André. 1952. Function, Structure, and Sound Change. *Word* 8: 1–32. [CrossRef]

Mayer, Mercer. 1969. *Frog, Where Are You?* New York: Dial Press.

McAuliffe, Michael, Michaela Socolof, Sarah Mihuc, Michael Wagner, and Morgan Sonderegger. 2017. Montreal Forced Aligner [Computer Program]. Version 0.9.0. Available online: http://montrealcorpustools.github.io/Montreal-Forced-Aligner/ (accessed on 17 January 2017).

McCloy, Daniel R. 2016. PhonR: Tools for Phoneticians and Phonologists. R Package Version 1.0-7. Available online: https://CRAN.R-project.org/package=phonR (accessed on 8 September 2020).

Mendoza-Denton, Norma. 1999. Fighting Words: Latina Girls, Gangs, and Language Attitudes. In *Speaking Chicana: Voice, Power, and Identity*. Edited by D. Letticia Galindo and María D. Gonzales. Tucson: University of Arizona Press, pp. 39–56.

Mendoza-Denton, Norma, and Melissa Iwai. 1993. They Speak more Caucasian: Generation Differences in the Speech of Japanese-Americans. In *Texas Linguistics Forum, Number 33*. Austin: University of Texas, Department of Linguistics, pp. 58–67.

Milroy, James, and Lesley Milroy. 1985. Linguistic Change, Social Network and Speaker Innovation. *Journal of Linguistics* 21: 339–84. [CrossRef]

Ohara, Yumiko. 1998. Two Languages, Two Cultures, and Two Vocal Apparati? Sociolinguistic Explanation for Phonetic Phenomena in Bilingual Speakers of Korean and Japanese. In *The Life of Language, the Language of Life: Selected Papers from the First College-Wide Conference for Students in Languages, Linguistics, and Literature*. Edited by Dina Yoshimi and Marilyn Plumlee. Honolulu: University of Hawaii at Manoa, pp. 124–33.

Parodi, Claudia. 2014. El español de Los Ángeles: Koineización y diglosia. In *Lenguas, Estructuras y Hablantes: Estudios en Homenaje a Thomas C. Smith Stark*. Edited by R. Barriga Villanueva and E. Herrera Zendejas. Tlalpan: El Colegio de México, Centro de Estudios Lingüísticos y Literarios, pp. 1101–23.

Pennock-Speck, Barry. 2005. The Changing Voice of Women. In *Actas del XXVIII Congreso Internacional de AEDEAN*. Edited by Juan José Calvo García de Leonardo, Jesús Tronch Pérez, Milagros del Saz Rubio, Carme Manuel Cuenca, Barry Pennock Speck and Maria José Coperías Aguilar. Valencia: Dept. de Filologia Anglesa i Alemanya, Univ. de València, pp. 407–15.

Pisanski, Katarzyna, Emanual C. Mora, Annette Pisanski, David Reby, Piotr Sorokowski, Tomasz Frackowiak, and David R. Feinberg. 2016. Volitional Exaggeration of Boday Size through Fundamental and Formant Frequency Modulation in Humans. *Scientific Reports* 6: 34389. [CrossRef]

Podesva, Robert J. 2011. The California Vowel Shift and Gay Identity. *American Speech* 86: 32–51. [CrossRef]

Podesva, Robert J., Annette D'Onofrio, Janneke Van Hofwegen, and Seung Kyung Kim. 2015. Country Ideology and the California Vowel Shift. *Language Variation and Change* 27: 157–86. [CrossRef]

Polinsky, Maria. 2018. *Heritage Languages and Their Speakers*. Cambridge: Cambridge University Press, vol. 159.

Pratt, Teresa, and Annette D'Onofrio. 2017. Jaw Setting and the California Vowel Shift in Parodic Performance. *Language in Society* 46: 283–312. [CrossRef]

Puzar, Aljosa, and Yewon Hong. 2018. Korean Cuties: Understanding Performed Winsomeness (Aegyo) in South Korea. *The Asia Pacific Journal of Anthropology* 19: 333–49. [CrossRef]

R Development Core Team. 2020. R Development Core Team, R: A Language and Environment for Statistical Computing. Available online: https://www.r-project.org/ (accessed on 8 September 2020).

Roeder, Rebecca V. 2010. Northern Cities Mexican American English: Vowel Production and Perception. *American Speech* 85: 163–84. [CrossRef]

Simpson, Adrian P. 2002. Gender-specific Articulatory-acoustic Relations in Vowel Sequences. *Journal of Phonetics* 30: 417–35. [CrossRef]

Sohn, Ho-min. 1999. *The Korean Language*. Cambridge: Cambridge University Press.

Stevens, Gillian. 1999. Age at Immigration and Second Language Proficiency among Foreign-born Adults. *Language in Society* 28: 555–78. [CrossRef]

Thomas, Erik R. 2000. Spectral Differences in/ai/Offsets Conditioned by Voicing of the Following Consonant. *Journal of Phonetics* 28: 1–25. [CrossRef]

Thomas, Erik R. 2001. An Acoustic Analysis of Vowel Variation in New World English. In *Publication of the American Dialect Society 85*. Durham: Duke University Press.

Traunmüller, Hartmut. 1997. *Auditory Scales of Frequency Representation*. Stockholm: Department of Linguistics, Stockholm University, Available online: http://www.ling.su.se/staff/hartmut/bark.htm (accessed on 8 September 2020).

Trofimovich, Pavel, Wendy Baker, and Molly Mack. 2011. Context- and Experience-based Effects on the Learning of Vowels in a Second Language. *Studies in the Linguistics Sciences* 31: 167–86.

Trudgill, Peter. 1972. Sex, Covert Prestige and Linguistic Change in the Urban British English of Norwich. *Language in Society* 1: 179–95. [CrossRef]

Tseng, Amelia. 2015. Vowel Variation, Style, and Identity Construction in the English of Latinos in Washington, D.C. Ph.D. dissertation, Georgetown University, Washington, DC, USA.

Tsukada, Kimiko, David Birdsong, Ellen Bialystok, Molly Mack, Hyekyung Sung, and James E. Flege. 2005. A Developmental Study of English Vowel Production and Perception by Native Korean Adults and Children. *Journal of Phonetics* 33: 263–90. [CrossRef]

Van Bezooijen, Reneé. 1995. Sociocultural Aspects of Pitch Differences between Japanese and Dutch Women. *Language and Speech* 38: 253–65. [CrossRef]

Wei, William. 1993. *The Asian American Movement*. Philadelphia: Temple University Press.

Weinfeld, Morton. 1985. Myth and Reality in the Canadian Mosaic: Affective ethnicity. In *Ethnicity and Ethnic Relations in Canada: A Book of Readings*, 2nd ed. Edited by Rita M. Bienvenue and Jay E. Goldstein. Toronto: Butterworths, pp. 65–86.

Werker, Janet F., and Richard C. Tees. 1984. Cross-language Speech Perception: Evidence for Perceptual Reorganization during the First Year of Life. *Infant Behavior and Development* 7: 49–63. [CrossRef]

Yang, Byunggon. 1996. A Comparative Study of American English and Korean Vowels Produced by Male and Female Speakers. *Journal of Phonetics* 24: 245–61. [CrossRef]

Yeni-Komshian, Grace H., James E. Flege, and Serena Liu. 2000. Pronunciation Proficiency in the First and Second Languages of Korean-English Bilinguals. *Bilingualism: Language and Cognition* 3: 131–49. [CrossRef]

Yoon, Kyuchul, and Soonok Kim. 2015. A Comparative Study on the Male and Female Vowel Formants of the Korean Corpus of Spontaneous Speech. *Phonetics and Speech Sciences* 7: 131–38. [CrossRef]

Yuasa, Ikuko Patricia. 2010. Creaky Voice: A New Feminine Voice Quality for Young Urban-oriented Upwardly Mobile American Women? *American Speech* 85: 315–37. [CrossRef]

Zampini, Mary. 2008. L2 Speech Production Research: Findings, Issues and Advances. In *Phonology and Second Language Acquisition*. Edited by Jette. G. Hansen Edwards and Mary Zampini. Amsterdam: John Benjamins, pp. 219–50.

languages

MDPI

Article

Interlingual Interactions Elicit Performance Mismatches Not "Compromise" Categories in Early Bilinguals: Evidence from Meta-Analysis and Coronal Stops

Joseph V. Casillas

Department of Spanish and Portuguse, Rutgers University, New Brunswick, NJ 08904, USA; joseph.casillas@rutgers.edu

Abstract: Previous studies attest that some early bilinguals produce the sounds of their languages in a manner that is characterized as "compromise" with regard to monolingual speakers. The present study uses meta-analytic techniques and coronal stop data from early bilinguals in order to assess this claim. The goal was to evaluate the cumulative evidence for "compromise" voice-onset time (VOT) in the speech of early bilinguals by providing a comprehensive assessment of the literature and presenting an acoustic analysis of coronal stops from early Spanish–English bilinguals. The studies were coded for linguistic and methodological features, as well as effect sizes, and then analyzed using a cross-classified Bayesian meta-analysis. The pooled effect for "compromise" VOT was negligible ($\beta = -0.13$). The acoustic analysis of the coronal stop data showed that the early Spanish–English bilinguals often produced Spanish and English targets with mismatched features from their other language. These performance mismatches presumably occurred as a result of interlingual interactions elicited by the experimental task. Taken together, the results suggest that early bilinguals do not have "compromise" VOT, though their speech involves dynamic phonetic interactions that can surface as performance mismatches during speech production.

Keywords: compromise VOT; voice timing; bilingualism; performance mismatches; dynamic phonetic interactions

Citation: Casillas, Joseph V. 2021. Interlingual Interactions Elicit Performance Mismatches Not "Compromise" Categories in Early Bilinguals: Evidence from Meta-Analysis and Coronal Stops. *Languages* 6: 9. https://doi.org/10.3390/languages6010009

Received: 28 October 2020
Accepted: 29 December 2020
Published: 4 January 2021

1. Introduction

Though early bilinguals with ample experience in their first (L1) and second (L2) languages are believed to show "monolingual-like" L2 speech production (Rao and Ronquest 2015), research on bilingual language modes (Grosjean 2001) has shown that cross-linguistic interactions are strengthened in bilingual contexts (Olson 2013; Simonet 2014) and can lead to production/perception that differs from that of monolinguals. A crucial question revolves around how a bilingual speaker mitigates producing/perceiving acoustically similar segments in their languages in unilingual and bilingual settings, i.e., whether the two systems are kept separate (see Magloire and Green 1999), or whether there is a compromise in the acoustic characteristics of the sounds of the two languages (see Caramazza et al. 1973). A body of research dating back to the 1970s shows that some early bilinguals produce the sounds of their languages in a manner that has been characterized as "compromise", "intermediate" or "merged" with regard to monolingual distributions (e.g., Flege and Hillenbrand 1984; Flege 1991; Flege and Eefting 1987b; Sundara et al. 2006, among many others). The present study takes a systematic look at the cumulative evidence for "compromise" speech production in early simultaneous and sequential bilinguals. This study begins by describing the nature of "compromise" categories and reviews the literature supporting their existence. Next, this study considers how "compromise" categories may arise under current models of bilingual phonology, and then describes some methodological concerns. Afterwards, this study employs meta-analytic techniques to assess the extant literature, and concludes by presenting an alternative account of "compromise" categories using production data from coronal stops.

1.1. Background and Motivation

A recurring finding in the bilingual speech literature is that even early bilinguals, both simultaneous and sequential, can display production, perception, and lexical processing that differs from monolinguals.[1] To be more specific, this finding is often framed in terms of "compromise" or "intermediate" phonetic categories in the bilingual production of stop contrasts, mainly because the acoustic properties of the segments are proposed to lie somewhere between those of the two languages. In the present work, a "compromise" category is defined as one that is not target-like with regard to some acoustic property. This particular line of research has focused on voice timing. Voice timing in stops can be determined by a number of parameters, the most common of which is voice-onset time (VOT, Lisker and Abramson 1964). VOT refers to the duration of the interval between the release of the stop and the onset of voicing of the following segment. VOT realizations include phonetically voiced stops, in which voicing begins before the release (i.e., lead VOT), as well as phonetically voiceless, lag stops, in which voicing begins after the release (i.e., short- and long-lag VOT, approx. 0–30 ms and 30+ ms, respectively). Many languages have two-way or three-way oppositions in which lead VOT, short-lag VOT, and long-lag VOT are mapped to phonologically voiced and voiceless categories. French, for example, contrasts voiced and voiceless stops with lead VOT and short-lag VOT, respectively. English, on the other hand, contrasts voiced and voiceless stops with short- and long-lag VOT, respectively.

The notion of "compromise" categories appears to have been coined by Williams (1980) in reference to Williams (1977), a study on Spanish–English bilinguals' production of bilabial stops. Williams (1980) stated, "The apparent development of compromise VOT targets in both perception and production for bilinguals and second-language learners may reflect a true convergence over time of the acoustic phonetic features of the two languages instead of the development of two separate phonetic systems" (p. 213). Subsequently, the term became commonplace in the bilingual speech literature (e.g., Flege and Hillenbrand 1984; Antoniou et al. 2010, 2011; Bullock et al. 2006; Chang et al. 2011; Flege 1991; Flege and Eefting 1987b, 1988; Gabriel et al. 2016; Jones 2020; Kehoe et al. 2004; Kiliç 2018; Kilpatrick 2004; Lein et al. 2016; Llama and López-Morelos 2016; López 2012; Morgan 2011; Sundara et al. 2006).

One explanation for the notion that bilinguals display "compromise" categories is offered by the Speech Learning Model (SLM, Flege 1995). The SLM posits that the ability to learn speech sounds is maintained throughout life. Novel phones are stored in long-term memory as phonetic categories. Importantly, L1 and L2 phonetic categories interact because they are assumed to share the same phonological space. L2 sounds are unconsciously and automatically linked with neighboring L1 phonetic categories via a mechanism Flege (1995) refers to as the "equivalence classification". According to the equivalence classification, linked phones that are perceived as being phonetically similar may result in a single, merged category used for both languages. If this is the case, the model posits that the L1 sound may assimilate to the L2 sound, resulting in both L1 and L2 productions that are intermediate in the phonetic space with regard to monolingual categories. If, on the other hand, the linked phones are perceived as being phonetically dissimilar, it is more likely that a new phonetic category will be formed. In this situation, the model predicts that the L2 category can dissimilate from nearby sounds in order to maintain phonetic contrast. Thus both scenarios can lead to "compromise" categories in bilingual speech.

Though they did not use the term "compromise" VOT, evidence for the phenomenon in early bilinguals dates back to a landmark study by Caramazza et al. (1973). This study examined voiceless stop production/perception in early French–English bilinguals that had acquired English before the age of seven. Caramazza et al. (1973) compared the bilinguals'

[1] The present work takes a broad view, similar to that of the studies referenced herein, on what constitutes early bilingualism. Specifically, an early bilingual is operationalized as one who was exposed to an additional language before the age of 12. This study also distinguishes between simultaneous and sequential bilingualism, where the former refers to an individual that acquires their languages at the same time and the latter refers to an individual that acquires an L2 after the L1.

performance on a perceptual identification task and a reading task with monolingual English and French speakers and found that, when the bilinguals were in French mode, they produced French /ptk/ no differently than the French monolingual control group. However, when they were tested in English mode, they produced English /ptk/ with less aspiration than the monolinguals. Caramazza et al. (1973) concluded that French–English bilinguals were "more closely aligned" with the French monolinguals, presumably because they had learned English sequentially.

The results from Caramazza et al. (1973) suggest that sequential learners that consistently use both languages over a long period of time may still produce stops in a way that differs from monolingual speech. An SLM account for these data might propose that category formation was blocked because of the equivalence classification. This would imply that the phones were still linked, though the bilinguals' L1 categories did not differ from those of monolingual controls, and, therefore, did not assimilate or merge.

In another seminal study in this literature, Flege and Eefting (1987b), examined the production of Spanish stops in numerous groups of Puerto Rican Spanish–English sequential bilingual children and adults with different linguistic backgrounds. Spanish stops contrast lead VOT with short-lag VOT similar to French. The bilinguals were compared with age-matched monolingual controls for both languages. Of particular relevance to the present work are two early bilingual groups that Flege and Eefting (1987b) referred to as earlier childhood bilinguals (ECB) and later childhood bilinguals (LCB). Both the ECB and LCB groups comprised adults who had learned English before the age of seven, but the ECB group was born in the U.S. mainland, or moved there shortly after birth, whereas the LCG group still lived in Puerto Rico. Both groups produced English /ptk/ with less aspiration than monolingual English speakers. Flege and Eefting (1987b) concluded that the early bilinguals—even those living in the U.S. since early childhood—were not able to produce English /ptk/ "authentically", suggesting that their "intermediate" productions lent support to the equivalence classification hypothesis. In other words, they continued to associate the Spanish and English stops as realizations of the same phonetic category.

The results from Caramazza et al. (1973) and Flege and Eefting (1987b) suggest that sequential learners will differ from monolinguals because of the equivalence classification, but "compromise" VOT has also been documented in simultaneous bilinguals (see Sundara et al. 2006; Fowler et al. 2008; Kupisch and Lleó 2017; Lein et al. 2016, among others). For instance, in a more recent study, Fowler et al. (2008) analyzed the voiceless stops of English–French simultaneous and early, sequential bilinguals by comparing their production with monolingual controls. The simultaneous bilinguals produced French /ptk/ with longer VOT than monolingual French speakers, and English /ptk/ with shorter VOT than monolingual English speakers. The effects were greater in the early, sequential bilinguals who produced stops that were characterized as even more intermediate in both languages. Though the SLM was designed with sequential learners in mind, Fowler et al. (2008) proposed that, for the simultaneous bilinguals, the different phones were not merged but still "cognitively identified with one another" because of the equivalence classification (p. 650).

The "compromise" VOT literature also shows significant between-study variability, as there are also studies that do not find "compromise" categories in bilingual production. For instance, Flege (1991) analyzed early and late Spanish–English bilinguals' production of Spanish and English /t/ in utterance initial and utterance medial position. Importantly, the voiceless coronal is produced with short-lag VOT in Spanish and long-lag VOT in English. Flege (1991) found that, in both positions, the late bilinguals produced English /t/ with "compromise" VOT, but the early bilinguals' production was no different from the monolingual speakers. Thus, it seems that some early bilinguals are able to establish se-parate phonetic categories and maintain separation between their languages, even when the segments are acoustically similar. Nevertheless, bilinguals do indeed seem to display more variability in their production of stops. For example, a recurring finding is that bilinguals whose L1 is a 'true voicing' language, that is, one that contrasts pre-voiced stops with short-lag stops (i.e., Spanish, French), tend to produce English voiced stops with pre-voicing at higher rates than monolingual English speakers (e.g., MacLeod and Stoel-Gammon 2005; Flege and Eefting 1987b; Hazan and Boulakia 1993). Much like in monolingual stop production (Chodroff and Wilson 2017), bilingual stop production also shows structured co-variation between stop categories (Chodroff and Baese-Berk 2019). This raises the possibility that variable voiced stop production might also be structured in other ways, on dimensions other than VOT.

1.2. Motivating Meta-Analysis

Given the discrepancies in the literature regarding the reliability of "compromise" categories, it is imperative that one consider all possible alternative explanations. One notable issue regarding voice timing that bares on this line of research is related to the mea-sure of VOT itself. Manifold studies show that VOT is modulated by linguistic factors, such as place of articulation (Cho and Ladefog 1999), word position (Antoniou et al. 2010), lexical stress (Casillas et al. 2015), and speech rate (Magloire and Green 1999) in monolingual and bilingual speech. For instance, faster speech is associated with shorter VOT and slower speech is associated with longer VOT, though the size of the effect may be language specific. Unfortunately, the large majority of studies related to stop contrasts in bilinguals do not take speech rate into account. One method to do so, proposed by (Stölten et al. 2015), is to use a relative measure of VOT by calculating the proportion of the duration VOT occupies in the stop + vowel sequence. This measure, relative VOT, has proven to be a more accurate, fine-grained metric that controls for possible speech rate confounds and provides more precise between-group comparisons.

In addition, bilingual language modes (Grosjean 2001) represent another factor that must be considered. Bilingual production/perception can vary (1) according to the mode (unilingual, bilingual) in which it is tested (e.g., Antoniou et al. 2010; Gonzales and Lotto 2013), and (2) as a function of the expectations the bilingual has about the communicative context (e.g., Gonzales et al. 2019; Lozano-Argüelles et al. 2020; Yazawa et al. 2019). These facts underscore the need to take special care when designing experiments so as to avoid confounds related to language modes.

Further methodological concerns include statistical power and sample size. Most studies in the social sciences test for small effects (Ellis 2010; cf. Plonsky and Oswald 2014), include small samples sizes, and, therefore, are underpowered (see Brysbaert 2020). The importance of this fact should not be overlooked, as an underpowered study is more likely to commit a type II error (false negative), and contributes to a literature with lower positive predictive value. Consequently, this implies that the prevalence of significant findings related to "compromise" categories may be indicative of publication bias.

Finally, advances in computational power have also led to more robust analytic techniques at the disposal of the speech researcher. Consider, for instance, multilevel modeling. These models provide two clear advantages: (1) they allow for partially pooled parameter estimates that are less affected by influential data points, and (2) they obviate the need to pool over subject and item repetitions in repeated measures designs. The combination of (1) and (2) helps to evade pseudoreplication in the phonetic sciences and reduces the likelihood of committing type II errors (Winter 2011).

1.3. The Present Study

The body of evidence suggesting that "compromise" categories in bilingual stop production may be fraught with confounds related to the primary outcome measure, as well as methodological issues related to language modes, power, sample size, and analytic techniques. This raises the question as to whether or not variable bilingual stop production may be best accounted for by some other phenomenon, such as dynamic phonetic interactions associated with language activation, rather than "compromised" underlying representations. All of the above motivate the need to assess the cumulative evidence via meta-analytic techniques. The present study aimed to address this need. Meta-analysis offers a principled method for assessing a body of research by using independent observations to derive an average effect size and, thus, draw an overall conclusion regarding the direction and magnitude of real-world effects. The "compromise" category literature would particularly benefit from meta-analysis because a large amount of research and theory building has been based on the early findings.

In order to assess the "compromise" category literature, the present study addressed the following questions:

1. What is the cumulative evidence for "compromise" categories?
2. Is the effect modulated by linguistic factors, such as place of articulation, word position, or lexical stress?
3. Do analytic techniques account for between-study variability?
4. Is the "compromise" category literature sufficiently powered?

In what follows, the present project responds to the aforementioned research questions by presenting a meta-analysis of the "compromise" VOT literature. Subsequently, this study provides an alternative account to the notion of "compromise" categories in early bilinguals using data from coronal stops.

2. Meta-Analysis

2.1. Method

2.1.1. Study Identification and Screening

The analysis employed a variety of techniques to locate relevant primary studies, focusing first on amassing a large study pool and later filtering out redundant or unusable records. The first step included searching library-housed online databases using various combinations of relevant keywords. The terms 'compromise categories', 'merged categories', 'mixed categories', and 'intermediate categories' were searched individually and in combination with 'early learners', 'early bilinguals' or 'simultaneous bilinguals', as well as 'VOT' or 'voice-onset time'. The databases included ERIC, Science Direct, Linguistics and Language Behavior Abstracts, PsycINFO, ProQuest Dissertations and Theses, and FirstSearch (see the supplementary materials for more details on the results from each search). Ancestry studies and studies citing the primary studies were also obtained via searches in Google and Google scholar. When potential studies were not available through the aforementioned resources, authors were contacted directly. There were 153,860 re-cords identified through database searching and 27 additional ancestry studies identified through Google and Google scholar. After removing duplicates and irrelevant hits, the study pool contained 148 records.

2.1.2. Eligibility

The 148 full-text articles and dissertations were assessed for eligibility. To be eligible, a study had to (1) include simultaneous and/or early bilinguals (AOA before 12 years old) that were adults at the time of testing, (2) examine a language pair with a two-way stop voicing contrast, specifically voiceless stops, and (3) include a monolingual comparison group.[2] For all studies, data from both languages were included in the meta-analysis if the participants were simultaneous bilinguals and control comparisons were included. For sequential learners, their second, sequentially learned language was utilized in comparison with controls, as "compromise" phonetic categories are more common in the literature in the L2. Languages with three-way contrasts (namely Korean) were excluded because these contrasts typically involve other parameters, i.e., pitch (see Holliday 2015). Finally, the study pool was limited to analyses of voiceless stops due to the fact that the majority of the records identified involved English, which allows both short-lag and lead VOT for phonologically voiced stops.

[2] An anonymous reviewer duly notes the fact that there is currently debate surrounding the use of monolingual control populations in bilingual research (see Sakai 2018). The present meta-analysis included monolingual controls as part of the search criteria because it best reflects the practices of research on "compromise" categories and thus led to the largest number of potential studies for the dataset.

The assessment resulted in a dataset of 68 studies that appeared to meet the aforementioned inclusion criteria. The studies spanned 6 decades, from the 1970s to the present, and came from a variety of sources, including journal articles, book chapters, MA and PhD theses, conference proceedings, and unpublished manuscripts. Forty-eight studies were discarded for a number of reasons: (1) they looked at something different (k = 15), (2) they did not include a control group for comparison (k = 16), (3) there was missing data (k = 10), (4) this study included duplicate data presented in a prior study (k = 2), or (4) this study examined a three-way contrast (k = 5). Requests for missing or unreported data were sent via email. One response out of 10 requests was received (with data). The search process led to a final dataset comprising 20 studies with 37 independent comparisons and a pooled participant sample size of 641.The average age cut-off used to classify participants as early bilinguals was 4.53 $\pm$ 2.84 SD. The majority of the usable studies were journal articles (n = 16), followed by MA/PhD theses and conference proceedings (n = 4). The usable studies spanned 5 decades, with 3 from the 1980s, 4 from the 1990s, 4 from the 2000s, 8 from the 2010s, and 1 study from 2020.

2.1.3. Coding

Each study was coded for linguistic and methodological features and effect sizes. The effect size was a measure of standardized mean difference (SMD), specifically Hedge's g. The linguistic features included the stop category (/p/, /t/, or /k/), lexical stress (stressed or unstressed syllable), and word position (initial, medial). The methodological features included analytic strategy (*t*-test, ANOVA, or LME) and pooling method for stops (individual evaluations of each segment versus averaging over combinations of segments). Effect sizes were calculated primarily using reported means and standard deviations (or standard errors).[3] Ultimately, word position was excluded as a moderator, as there were not enough studies that included this factor.

2.1.4. Statistical Analysis

A cross-classified Bayesian meta-analysis was conducted by fitting the study data with the multilevel regression model formulated below:

$$\begin{aligned} \text{SMD}_i &\sim \text{Normal}(\theta_i, \sigma_i = \text{se}_i) \\ \theta_i &\sim \text{Normal}(\mu, \tau) \\ \mu &\sim \text{Normal}(0, 1) \\ \tau &\sim \text{HalfCauchy}(0, 1) \end{aligned} \tag{1}$$

Effect size (SMD) was the outcome variable and lexical stress (stressed, unstressed) and analytic strategy (LME, other) were included as population-level effects (i.e., fixed effects). The likelihood of the outcome variable was assumed to be a gaussian distribution. Individual studies and stop pooling methods were group-level effects (i.e., random effects). Population-level effects were deviation coded (lexical stress: stressed = 0.5, unstressed = -0.5; analytic strategy: LME = 0.5, other = -0.5) such that the posterior distribution of model estimates for each effect provided an assessment of effect size. The model included regularizing, weakly informative priors (Gelman et al. 2017) which were normally distributed and centered at 0 with a standard deviation of 1 for all population-level parameters. A cauchy prior set at 0 with scale 1 was used for τ. Fianlly, the model was fit with 4000 iterations (2000 warm-up) and Hamiltonian Monte-Carlo sampling was carried out with 4 chains distributed across 4 processing cores. The analysis was conducted in R (R Core Team 2019, version 4.0.3) and was fit using stan (Stan Development Team 2018) via the

[3] In the case of one study, effect size was calculated from the reported degrees of freedom and F-value of a one-way ANOVA. For three studies standard deviations were not reported but the manuscript included boxplots. In these cases, the median and interquartile range were derived from the boxplots via webplot digitizer (Rohatgi 2020) and then used to calculate the mean and standard deviations (see Wan et al. 2014). All figures used for these approximations are available in the supplementary materials.

R package brms (Bürkner 2017). More information regarding Bayesian Data Analysis is available in the supplementary materials. All supplementary analyses as well as the data, code, and the experimental materials necessary to reproduce the analyses reported in this article are available at: https://osf.io/un45x/.

2.2. Results

Summaries of the posterior distribution of the meta-analytic model are provided in Figures 1 and 2 (see supplementary materials for summaries in table format). Averaging over lexical stress and analytic strategies, the pooled estimate of the standardized mean difference (SMD) was small and negative ($\beta = -0.132$, HDI = $[-0.708, 0.468]$). As reflected in Figure 1A, the posterior distribution of the effect size estimate is wide and encompasses plausible values on both sides of a point null of 0. In short, there is not compelling evidence in support of a difference in VOT between bilinguals and monolingual controls. The estimates of variability from group-level effects *pooling method* and *individual studies* are plotted in Figure 1B. One can see that *individual studies* were a considerable source of variability. Figures 1C and 2 illustrate estimate uncertainty in comparison with the overall pooled effect. The moderator (subgroup) effects were negligible. Specifically, lexical stress had no effect on group differences ($\beta = -0.14$, HDI = $[-0.766, 0.483]$), nor did analytic strategy ($\beta = 0.182$, HDI = $[-0.657, 1.035]$), though mixed effects models tended to narrow the gap between group difference estimates, as illustrated by the positive β-parameter (see Figure 1D).

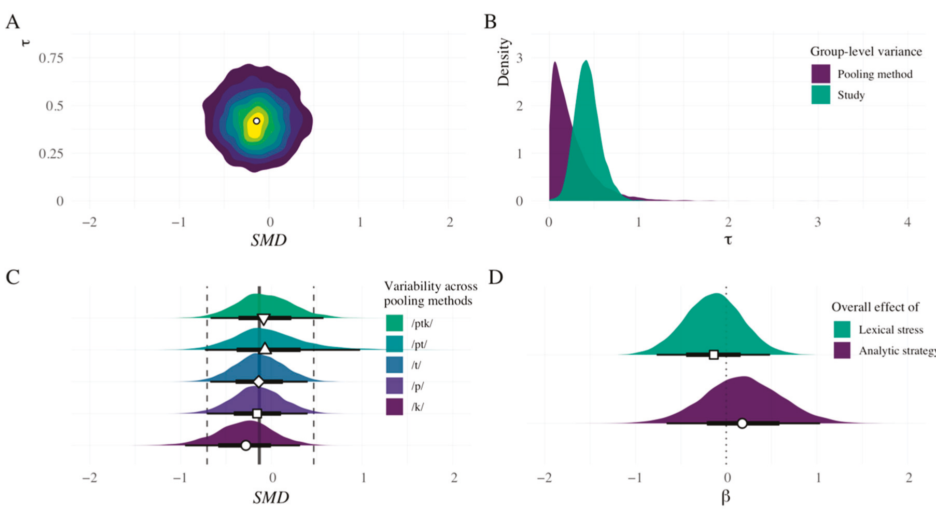


Figure 1. Summary of the posterior distribution of the meta-analytic model. Panel (**A**) plots the pooled estimate of effect size (SMD) on the horizontal axis and standard error (τ) on the vertical axis. Lighter colors illustrate higher density areas (i.e., more plausible values) and the point represents the posterior median. Panel (**B**) illustrates estimates of variance from group-level effects and Panel (**C**) provides a sub-category summary of variability as a function of pooling method. The vertical lines show the posterior median (solid) and 95% credible interval (dashed) of the pooled effect. Panel (**D**) summarizes the posterior distributions of the overall effects of the lexical stress and analytic strategy moderators.

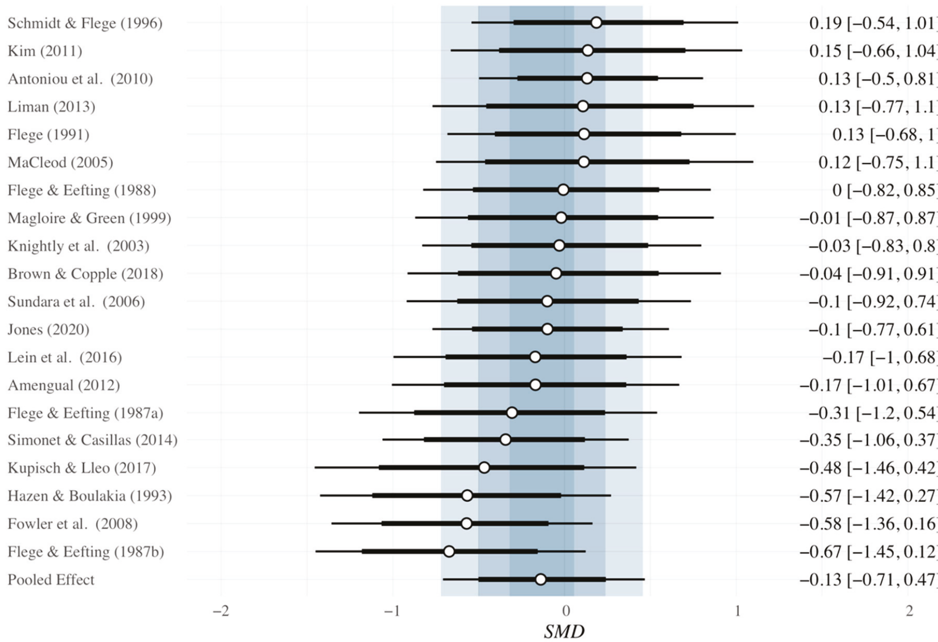


Figure 2. Summary of posterior model estimates for individual studies (vertical axis) and the effect size (SMD, horizontal axis). White points represent posterior medians and horizontal bars capture the ± 95% and 80% credible intervals, which are also printed along the right vertical margin. The vertical shaded rectangle illustrates ± 95%, 80%, and 50% credible intervals around the posterior median of the pooled effect.

2.3. Interim Discussion

Using a variety of techniques to locate relevant research on bilingual stop production, a dataset of 20 studies with 37 independent, usable bilingual-to-monolingual comparisons was collected. The studies were coded for linguistic and methodological features and effect sizes were calculated in order to estimate the cumulative effect in the literature comparing bilingual voiceless stop production to that of monolingual controls via meta-analysis. The results of the meta-analysis suggest that the cumulative effect in the literature is negligible and includes a high degree of uncertainty. The pooled estimate of the present dataset is −0.132 standard deviations, 95% HDI = [−0.708, 0.468]. Traditional standards classify an effect size of 0.20 as small, 0.50 as medium, and 0.80 as large (see Cohen 2013; Ellis 2010). Plonsky and Oswald (2014) suggest even more stringent standards for L2 research (small: d = 0.40, medium: d = 0.70, large: d = 1.00). In this dataset, the posterior probability that the average effect size meets or exceeds Plonsky and Oswald (2014) suggestion for a small effect is 0.18. Why, then, has a significant amount of the literature built theory around the "compromise" category claim?

One possibility is publication bias. Figure 3 provides a funnel plot of the unpooled dataset comparing standard error as a function of effect size (SMD). If the literature on bilingual stop production suffered from publication bias one would expect to see individual studies (points) dispersed asymmetrically around the pooled estimate. This does not appear to be the case, as one observes studies on both sides of the pooled estimate (the vertical black line).

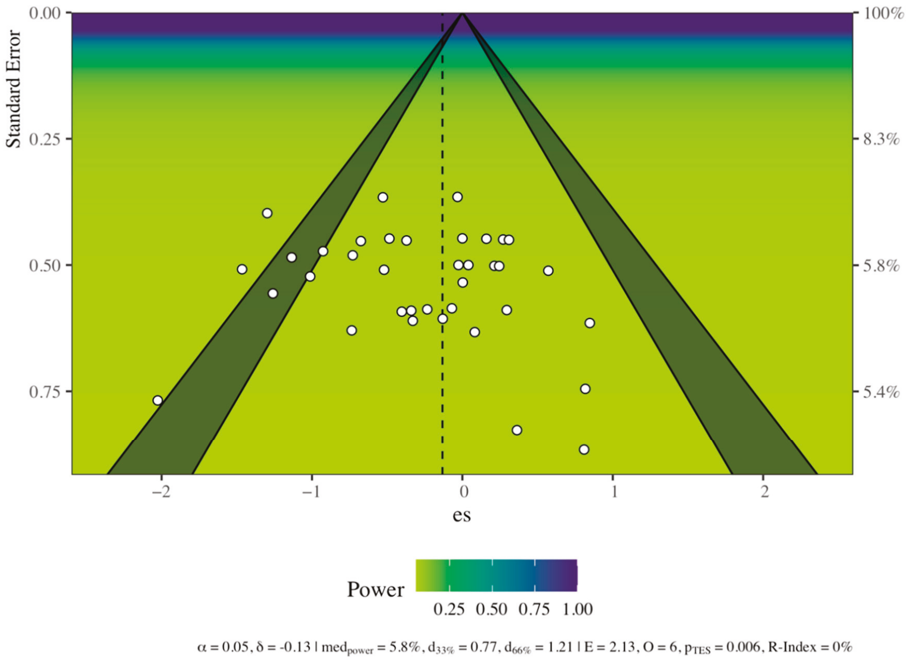


Figure 3. Funnel plot of effect sizes (Hedge's g, horizontal axis), standard error, and power (vertical axis). Power for individual studies was calculated using the pooled estimate (−0.132 SMD, vertical dashed line) from the meta-analytic model. Background color illustrates power (darker colors represent more power).

If we were to assume that the "compromise" category effect in bilingual speech were real, then we are left questioning why it was not borne out in the meta-analysis. One possibility, a glaring shortcoming of the extant literature, is the fact that all of the studies analyzed are underpowered (median = 5.80%). The gold standard for power suggested in psychology is 80% (see Ellis 2010), and none of the studies included in the present analysis met this standard (see right margin of Figure 3). The lack of power in this literature likely results from small sample sizes. If we again assume that "compromise" categories exist in bilingual speech and that the effect is small—0.40 using Plonsky and Oswald's suggestion for L2 research—a hypothetical study would need to include 99 participants *per group* to have an 80% chance of capturing the effect with alpha set a 0.05. The number is even more astounding if we use the pooled estimate of the present study and assume the same conditions: 896 participants, again, *per group*.

Another possible factor contributing to the low overall power in the extant literature is measurement error. While measuring VOT is rather straightforward, the majority of the early research in this area required researchers to average VOT values for individuals over items, as well as over repetitions of items. This suggests that the values being used in statistical analyses may inherently misrepresent the actual productions of the participants. If, for example, a participant produces the voiced stops of English variably, with and without pre-voicing, a mean value for this participant might not accurately represent a prototypical stop from either distribution. Given the structured variability observed in monolingual (Chodroff and Wilson 2017) and bilingual (Chodroff and Baese-Berk 2019) stop production, it seems reasonable to assume that there is also inherent structure manifested in other ways. Advancing technology has made the costly computations involved with partial pooling methods more accessible to researchers, and, as a result, this is reflected in more recent studies using more powerful analytic strategies, such as multilevel models. The present meta-analysis took analytic strategies into account and found no effect, though this may be explained by the fact that the majority of the studies included did not use partial pooling (k = 33 out of 37 independent comparisons).

It is important to note that the meta-analysis presented here is limited in several non-trivial ways. First, the pool of studies excluded clearly relevant, seminal research due to uncontrollable circumstances (e.g., missing data). Second, the criteria for inclusion necessarily limited the sample to voiceless stop production from early/simultaneous bilinguals speaking languages with two-way contrasts. While this filtering facilitated controlling possibly confounding factors, it raises questions regarding what data from other sources can contribute to the cumulative research on "compromise" categories in bilinguals. If, for example, "compromise" categories are not real in early and simultaneous bilinguals, what, then, can explain the observed variability in bilingual stop production? The following section considers coronal stop data from highly proficient Spanish–English early bilinguals in order to explore the interaction between language-specific voice-timing and place of articulation differences when both languages are activated in bilingual mode.

3. Production of Coronal Stops

This section presents an acoustic analysis of coronal stop data from early Spanish–English bilinguals in order to explore alternative explanations for "compromise" categories. The bilinguals completed a delayed shadowing task in which both languages were highly activated in order to elicit bilingual language mode. While English contrasts /d t/ through short-lag VOT and long-lag VOT, respectively, Spanish has the same voicing distinction through lead VOT and short-lag VOT. Importantly, the coronal stops of each language also differ regarding place of articulation. In Spanish coronal stops are described as dental, whereas in English they are described as alveolar (Casillas et al. 2015).

3.1. Method

3.1.1. Participants

The dataset included 33 participants between the ages of 18 and 23, all of which were female. There were 17 bilingual participants and 16 monolingual controls. Of the monolingual participants, eight were native Spanish speakers, born and raised on the island of Majorca, Spain. The remaining eight were native English speakers, born and raised in the US Southwest.

The Spanish–English bilinguals came from Southern Arizona and Northern Mexico. They were raised in Spanish-speaking families and were schooled mostly in English in Southern Arizona. They reported using English and Spanish daily, both in the classroom as well as with their friends and relatives. The bilingual group completed the Bilingual Language Profile (BLP, Gertken et al. 2014) in order to assess language dominance. The BLP calculates a weighted average of language dominance based on the individual history, use, proficiency, and attitudes of the bilinguals with regard to their languages. The measure ranges from −218 to 218 with values near the extremes implying dominance in one of the languages. Values close to 0 are taken as an indication of balanced bilingualism. In the present study, Spanish was arbitrarily assigned to positive values. Figure 4 plots language dominance (Panel A) and language use and proficiency data (Panel B) derived from the BLP. The bilingual group had a mean dominance score of 2.08 (SD = 40.42), suggesting rather balanced bilingualism (Panel A). Participants that reported using Spanish more often also tended to report being more proficient in that language; the converse was also true for English (Panel B).

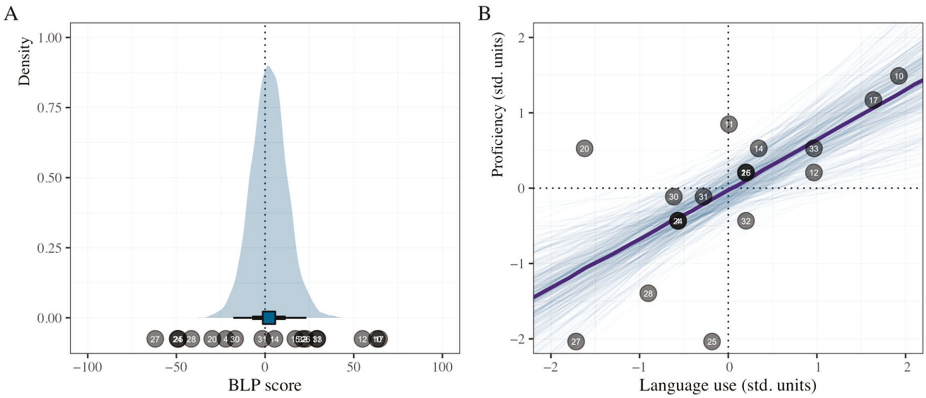


Figure 4. Bilingual Language Profile data. Panel (**A**) illustrates the distribution of BLP dominance scores. Panel (**B**) shows language proficiency as a function of language use. Both measures come from self-report data and are plotted in standardized units.

3.1.2. Materials

There were 48 target words (English: k = 24; Spanish: k = 24) that contained voiced and voiceless coronal stops in word initial position. For each language, 12 targets began with /d/ and 12 began with /t/, equally divided between stressed and unstressed syllables (see supplementary materials). All stops were followed by a low vowel (/a/ for Spanish and /æ, ɑ/ for English).

The participants completed a delayed repetition task in which they heard the target words presented in a carrier phrase ("x is the word" or the Spanish equivalent "x es la palabra"). The auditory stimuli were recordings of six male native speakers: three native English speakers and three native Spanish speakers. These recordings served as the auditory stimuli repeated out loud by the participants in the delayed repetition task. Words not containing coronal stops were considered distractors (k = 20). Praat was used to present the sentences randomly in auditory form and the speakers were asked to listen to the entire sentence and then repeat it out loud at their own pace.

The monolingual English speakers and bilinguals were recorded in a sound attenuated booth. The monolingual Spanish speakers were recorded in a quiet classroom on the campus of the *Universitat de les Illes Balears* in Majorca, Spain. The monolingual English speakers were recorded in English and the monolingual Spanish speakers were recorded in Spanish. The Spanish–English bilinguals were recorded in both of their languages in a single session with all English and Spanish items presented in a single, randomized block in order to activate both languages. The full dataset included 3519 tokens (24 target words per language × 3 repetitions). Eighty-one items (2.25%) were discarded due to mispronunciations or extraneous noise. A Shure SM10A dynamic head-mounted microphone with a Sound Devices MM-1 microphone pre-amplifier captured the acoustic signal and it was saved to a Marantz PMD660 digital speech recorder. The signal was digitized at 44.1 kHz and 16-bit quantization.

3.1.3. Measurements

The audio files were low-pass filtered at 11.025 kHz. Synchronized waveform and spectrographic displays were used to mark the onset of modal voicing and of the stop burst, as well as the offset of the first vowel. The onset of voicing was operationalized as the upwards zero-crossing of the first periodic pattern in the oscillogram and the offset of the vowel was marked at the downwards zero-crossing of the final periodic pattern. VOT was calculated as the difference (in ms) between the onset of modal voicing and the onset of the burst. Relative VOT was calculated as the ratio between VOT and the total duration of the stop-vowel sequence. Spectral moment measures were calculated from a 6 ms window beginning at the onset of the burst. Specifically, kurtosis was extracted from the spectral envelope, which ranged from 60 Hz to 11.025 kHz.

3.1.4. Statistical Analysis

The bilingual voice-timing data were analyzed using Bayesian multilevel regression models. Specifically, separate models were fit for VOT and relative VOT as a function of voicing (voiced, voiceless) and language (English, Spanish). Fixed effects were deviation coded (voicing: voiced = 0.5, voiceless = −0.5; language: English = 0.5, Spanish = −0.5) thus the posterior distribution provided an assessment of effect size for each predictor. Item repetition was included as a continuous predictor and was centered to have a mean of 0. The random effects structure included a by-subject intercept with a random slope for voicing, as well as a by-item intercept. The model included regularizing, weakly informative priors (Gelman et al. 2017) which were normally distributed and centered at 0 for all population level parameters. The standard deviation was set at 50 and 3 for the VOT and relative VOT models, respectively. The models were fit with 4000 iterations (1000 warm-up) and Hamiltonian Monte-Carlo sampling was carried out with 4 chains distributed across 4 processing cores. For each model a region of practical equivalence (ROPE) was established around a point null value of 0 (see Kruschke 2018) using the following formula:

$$\frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2+\sigma_2^2}{2}}} \tag{2}$$

For all models, median posterior point estimates are reported for each parameter of interest, along with the 95% highest density interval (HDI), the percent of the region of the HDI contained within the ROPE, and the maximum probability of effect (MPE). For statistical inferences, a posterior distribution for a parameter β in which 95% of the HDI falls outside the ROPE and a high MPE (i.e., values close to 1) were taken as compelling evidence for a given effect. Again, the analyses were conducted in R (R Core Team 2019, version 4.0.3) and models were fit using stan (Stan Development Team 2018) via the R package brms (Bürkner 2017).

3.2. Results

Figure 5 plots the VOT data as a function of group and stop category. Looking across the horizontal axis, one can observe distinct distributions for voiced and voiceless stops. In comparing the three groups along the vertically faceted panels, it becomes particularly clear that the bilinguals produce the coronal stops similarly to the monolingual controls in English and Spanish; however, upon close inspection, we can see that the bilingual group produces more pre-voiced /d/ tokens in English than the monolingual English speakers, as well as more short-lag /d/ tokens than monolingual Spanish speakers. When pooled together to calculate by-subject averages, tokens such as these could skew the measurement, fostering the notion that the bilinguals produce some segments with intermediate values that do not correspond with prototypical monolingual values of either language. The analysis of the voice-timing data did not provide compelling evidence that the bilingual productions differed from monolingual controls.

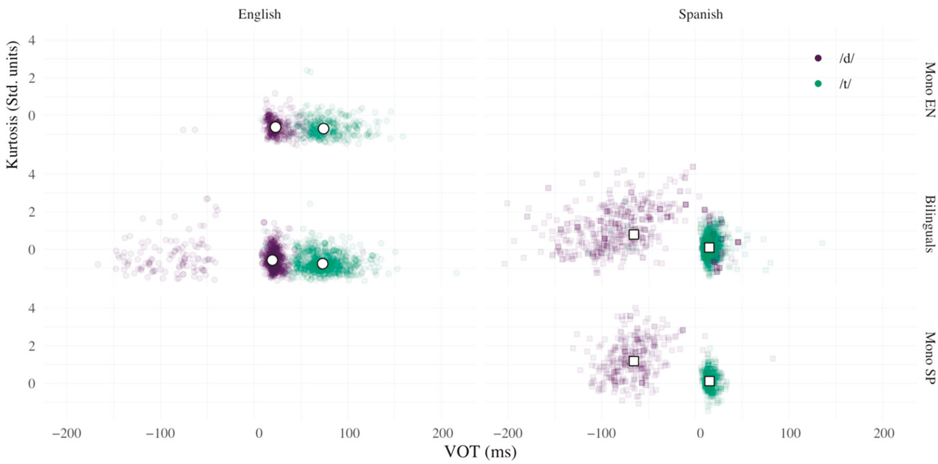


Figure 5. Voice-onset time (horizontal axis) and Kurtosis (vertical axis) as a function of speaker group and stop category. Colored points represent raw data. White points represent category medians.

3.2.1. VOT

For Spanish, phonologically voiced stops had shorter VOT than voiceless stops (β = −38.865, HDI = [−46.555, −31.766], ROPE = 0, MPE = 1), but VOT did not vary as a function of group (β = 0.78, HDI = [−6.977, 8.604], ROPE = 0.844, MPE = 0.579), nor did the two factors interact (β = 1.213, HDI = [−4.842, 8.058], ROPE = 0.898, MPE = 0.646). For English stops, voiced stops also had shorter VOT than voiceless stops (β = −31.743, HDI = [−37.77, −25.363], ROPE = 0, MPE = 1). Averaging over stop category, the analysis suggested that bilingual VOT was slightly shorter than that of the monolingual English speakers (β = −7.005, HDI = [−13.015, −0.951], ROPE = 0.19, MPE = 0.988), though approximately 19% of the most plausible estimates fell within our *a priori* established region of practical equivalence. There was no evidence of a group × phoneme interaction (β = −5.161, HDI = [−11.02, 0.18], ROPE = 0.41, MPE = 0.966).

3.2.2. Relative VOT

The more stringent relative VOT metric tells a similar story. For Spanish, voiced stops comprised a larger proportion of the stop-vowel sequence than voiceless stops (β = 0.1, HDI = [0.071, 0.127], ROPE = 0, MPE = 1), but the stop-vowel ratio did not vary between groups (β = −0.001, HDI = [−0.018, 0.015], ROPE = 0.945, MPE = 0.576), nor was there a two-way interaction (β = −0.004, HDI = [−0.028, 0.017], ROPE = 0.792, MPE = 0.658). For English, voiced stops comprised a smaller proportion of the stop-vowel sequence than voiceless stops did (β = −0.101, HDI = [−0.119, −0.083], ROPE = 0, MPE = 1). Averaging over stop category, there was no difference between groups (β = 0.004, HDI = [−0.013, 0.02], ROPE = 0.898, MPE = 0.697), and, finally, the two factors did not interact (β = 0.01, HDI = [−0.004, 0.025], ROPE = 0.699, MPE = 0.911).

The complete summary of the posterior distributions of all models are available in Table 1. As a point of comparison, the VOT and relative VOT data were also fit using a 2 × 2 repeated measures ANOVA under a null-hypothesis significance testing (NHST) frequentist framework by averaging over items and item repetitions. The VOT model, and not the relative VOT model, suggested a group × stop category interaction for the English data. The complete model summaries for all analyses are available in the supplementary materials.

Table 1. Summary of the posterior distribution modeling voiceless responses as a function of VOT, context, z-LexTALE, and order. The table includes posterior means, the 95% HDI, the percentage of the HDI within the ROPE, and the maximum probability of effect (MPE).

Outcome	Language	Parameter	β	95% HDI	MPE	ROPE %	ROPE
VOT	Spanish	Intercept	−23.19	[−31.84, −14.89]	1	0	[−5.12, 5.12]
		Group	0.78	[−6.98, 8.60]	0.58	0.84	[−5.12, 5.12]
		Phon.	−38.87	[−46.56, −31.77]	1	0	[−5.12, 5.12]
		Item rep.	−1.02	[−2.73, 0.68]	0.89	1	[−5.12, 5.12]
		Group × Phon.	1.21	[−4.84, 8.06]	0.65	0.89	[−5.12, 5.12]
	English	Intercept	45.19	[38.66, 51.81]	1	0	[−4.60, 4.60]
		Group	−7.00	[−13.02, −0.95]	0.98	0.19	[−4.60, 4.60]
		Phon.	−31.74	[−37.77, −25.36]	1	0	[−4.60, 4.60]
		Item rep.	−0.21	[−1.57, 1.13]	0.62	1	[−4.60, 4.60]
		Group × Phon.	−5.16	[−11.02, 0.18]	0.96	0.41	[−4.60, 4.60]
Relative	Spanish	Intercept	0.21	[0.19, 0.23]	1	0	[−0.01, 0.01]
VOT		Group	−0.01	[−0.02, 0.02]	0.58	0.94	[−0.01, 0.01]
		Phon.	0.10	[0.07, 0.13]	1	0	[−0.01, 0.01]
		Item rep.	0.01	[−0.01, 0.01]	0.83	1	[−0.01, 0.01]
		Group × Phon.	−0.01	[−0.03, 0.02]	0.66	0.79	[−0.01, 0.01]
	English	Intercept	0.29	[0.27, 0.31]	1	0	[−0.01, 0.01]
		Group	0.01	[−0.01, 0.02]	0.69	0.89	[−0.01, 0.01]
		Phon.	−0.10	[−0.12, −0.08]	1	0	[−0.01, 0.01]
		Item rep.	−0.01	[−0.01, 0.01]	0.83	1	[−0.01, 0.01]
		Group × Phon.	0.01	[−0.01, 0.03]	0.91	0.69	[−0.01, 0.01]

3.2.3. Bilingual Performance Mismatches

Taken together, the aforementioned analyses do not provide compelling evidence that Spanish–English bilinguals produce Spanish and English coronal stops in a manner that robustly differs from their monolingual counterparts. That being said, there does appear to be a qualitative difference between the bilinguals and the monolingual controls in terms of variability. Specifically, on occasion, the bilinguals appear to produce English /d/ with pre-voicing, as well as Spanish /d/ and English /t/ with short-lag VOT. In order to analyze this further, the data were subset based on these mismatched VOT properties and the relationship between voice timing (relative VOT) and place of articulation (Kurtosis of the stop burst) was explored.

The scatter plots in Figure 6 show the mismatched targets for /d/ (Panel A) and /t/ (Panel B) as a function of target language. The vertical axis is kurtosis (standardized units) and the horizontal axis is relative VOT (standardized units). Higher kurtosis values are found in Spanish monolingual coronals with regard to English monolingual coronals. This difference reflects the place of articulation differences between Spanish (dental) and English (alveolar) coronals (see Casillas et al. 2015; Sundara et al. 2006).

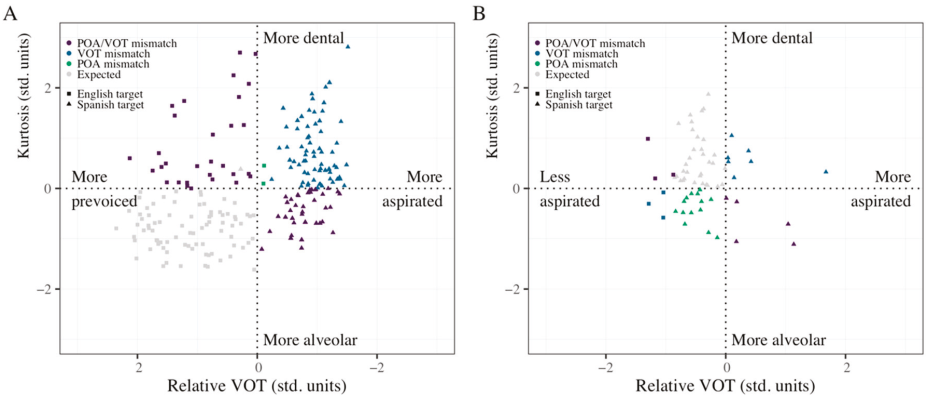


Figure 6. Summary of performance mismatches for /d/ targets (Panel **A**) and /t/ targets (Panel **B**) in bilinguals. In both panels, English targets are triangles and Spanish targets are squares. The horizontal axis represents relative VOT and the vertical axis represents kurtosis in standardized units.

The plots can be interpreted using the quadrants specified by the vertical and horizontal dotted lines. Points on the left side of the vertical line are associated with lower VOT values and points to the right side are associated with higher VOT values.[4] Points above the horizontal dotted line contain burst characteristics consistent with dental place of articulation (POA), while points below the horizontal line contain burst characteristics consistent with alveolar POA. The colors of the points are associated with the type of mismatch, which could be VOT, POA, or both VOT and POA (grey points are tokens produced as expected that met the filtering criteria). Of particular interest are the VOT/POA mismatches in the upper-left and lower-right quadrants of both plots.

The majority of the mismatched productions occurred in phonologically voiced target items (Panel A). One can observe Spanish targets that were realized as short-lag stops with more alveolar bursts (purple triangles), as well as English items that were produced with pre-voicing and more dental bursts (purple squares). The phonologically voiceless target items led to fewer mismatches (Panel B), though there are instances of English targets with short-lag VOT and more dental bursts, as well as Spanish targets with long-lag VOT and more alveolar bursts. The amount of mismatches produced was not associated with language dominance or any of the self-report measures collected in the BLP (see supplementary materials for more details).

4. Discussion

The present study included two primary analyses: a meta-analysis of "compromise" VOT and coronal stop data from a delayed shadowing production experiment. The results of the meta-analysis suggested that the pooled estimate of the cumulative effect-size was small, and just as likely to be positive as it was to be negative. The model considered linguistic factors as well as methodological factors. There was no evidence that lexical stress was a relevant moderator for "compromise" VOT, nor that analytic strategies resulted in a higher or lower likelihood of encountering differences between bilinguals and monolinguals. Between-study differences accounted for a larger proportion of the variance than pooling methods for the stop categories. The posterior estimate of the pooled effect made it particularly clear that the "compromise" VOT literature is underpowered (median = 5.80%).

[4] Note that the *x* axis was reversed in panel A of Figure 6 so that both panels are interpreted in the same manner, that is, with lower VOT values on the left and higher VOT values on the right. The reason the *x* axis must be reversed for this to be true is because pre-voiced stops have higher relative VOT values, i.e., they account for a larger portion of the stop + vowel sequence, than short-lag stops. The same is true for long-lag stops when compared with short-lag stops.

That being said, there is no evidence suggesting that this literature suffers from publication bias.

The analysis of the coronal stop data produced two findings. First, bilingual stop production is highly variable, and, second, this variability is structured in consistent, predictable ways. Specifically, the analysis showed that the bilingual speech contained target mismatches. That is, when producing English targets, the bilingual speech included pre-voiced /d/ tokens and short-lag /t/ tokens. When speaking Spanish, the bilingual speech included short-lag /d/ tokens. Crucially, these mismatches also displayed burst characteristics that were consistent with the mismatched language, suggesting that, on the whole, not only were the bilinguals using the voice timing of their other language, but also the corresponding place of articulation. The performance mismatches are attributed to the fact that the experiment induced high activation of both Spanish and English by being conducted in "bilingual" mode.

Taken together, the results of the present work depict bilingual stop production in a new light. Specifically, it seems unlikely that early bilinguals have "compromise" categories for stops, but rather produce performance category mismatches that result from dynamic phonetic interactions associated with language activation. This assertion is inconsistent with the evidence put forth by many studies in this literature. Possible explanations for the discrepancy may revolve around methodological issues. For example, the effects of bilingual language modes on speech production/perception are well attested. Many, but not all, of the earliest studies in the "compromise" VOT literature control for language mode by conducting experimental sessions on different days, with all materials and interactions with the experimenter conducted in the relevant language. Thus, it seems unlikely that language mode is a confounding factor in this body of literature. A more likely candidate is speech rate, which is negatively correlated with VOT (see Schmidt and Flege 1996; Magloire and Green 1999). Stölten et al. (2015) proposed controlling for speech rate by using relative VOT when making between group comparisons. Only two of the studies included in the present analysis controlled for speech rate. Magloire and Green (1999) found that Spanish–English bilinguals' English bilabial stop production was no different from that of monolingual English speakers in slow, normal and fast speech.[5]

The SLM posits that if learners can avoid the equivalence classification of two similar phones, they will be able to develop new phonetic categories. When this occurs, one prediction is that the L2 phonetic category will deflect or dissimilate from neighboring categories in order to maintain phonetic contrast. The early bilinguals of the present study clearly established a phonetic category for English /t/. In terms of the equivalence classification, it is more difficult to make a determination regarding /d/, which was realized with pre-voicing at a higher rate than in the monolingual controls. Any claim that increased pre-voicing is indicative of a "compromise" category would be weak, at best, as American English stops are often pre-voiced (Lisker and Abramson 1964) and may even be the default for some varieties (Walker 2020). Given the acoustics of the category mismatches, there does not appear to be any reason to believe that the bilinguals underlying phonetic categories are anything different from the input they receive in their speech community. On the contrary, the results herein support models that do not assert a bi-directional influence on the underlying grammar.

The Second Language Linguistic Perception Model (L2LP, Escudero 2005; Van Leussen and Escudero 2015), for example, proposes that the L2 grammar is separate from and develops independently of the L1 grammar. This model conceives of Grosjean's bilingual language modes as a continuum ranging from a unilingual mode in the L1 on one extreme to a unilingual mode in the L2 on the other, with an L1–L2 bilingual mode in the middle. Importantly, the L1 and L2 underlying grammars can be activated selectively or in parallel in real time. Language activation can be triggered by variables that are linguistic or

[5] Other studies on bilingual stop production have taken speech rate into account, to be sure, but were not included in the present analysis due to missing and/or unavailable data.

extralinguistic in nature, such as the use of cognates in the experimental items (e.g., Amengual 2012) or the participant's beliefs about the language required for a given task. Thus, this model can account for phenomena like the double phonemic boundary effect (e.g., Lozano-Argüelles et al. 2020) and language-dependent cue weighting (e.g., Yazawa et al. 2019) in speech perception.

To the author's best knowledge, the L2LP has not been used to account for production of consonants, but, on the surface, it appears to provide an elegant explanation for bilingual performance mismatches via parallel activation in bilingual mode. For instance, the bilinguals in the present study would have a monolingual Spanish mode on one end, where they produce /d t/ as dental stops with pre-voicing and short-lag VOT, respectively, and a monolingual English mode on the other, where /d t/ are realized at alveolar place with short- and long-lag VOT. When both languages are activated in parallel in bilingual mode, all combinations of place and voice settings are available, though not equally probable.

It is worth noting that neither the SLM nor the L2LP were designed to explain simultaneous bilingualism. Moving forward, a complete model of bilingual phonology should be able to account for behavior of both simultaneous and sequential bilinguals with a wide variety of linguistic experience in order to appropriately model the nature of the dynamic phonetic interactions that occur between robust phonetic sub-systems in diverse communicative contexts.

The present work could be improved by including more of the extant literature on "compromise" VOT. The meta-analysis presented here excluded a non-trivial subset of relevant studies due to missing data. As mentioned in the interim discussion, the whole of the "compromise" VOT literature includes small sample sizes and is underpowered. The coronal stop data presented herein is subject to the same critique. This analysis was necessarily exploratory in nature, serving the main purpose of providing a qualitative assessment of performance mismatches. Future studies should further examine the nature of performance mismatches to shed light on how they might be modulated by proficiency, language dominance, and language modes. The present analysis focused on stops, specifically coronal stops, and recent research proposes that alveolar sounds may enjoy a special status in L2 speech learning due to a universal phonetic bias (Bohn 2020). Thus, future research should focus on other speech segments, particularly those where multiple cues are weighted differently between language pairs, in order to better understand performance mismatches in conjunction with bilingual language modes. The aforementioned avenues of inquiry motivate testable hypotheses that could prove fruitful in the continued development of models of bilingual phonology.

5. Conclusions

The present study combined meta-analytic techniques and coronal stop data to assess the extent to which early bilinguals have "compromise" categories for voiceless stops. The results of the analyses (1) suggest that there is little evidence to support this claim in the extant literature, and (2) reinforce the notion that bilingual speech involves dynamic phonetic interactions that can surface as performance mismatches during speech production. The data provide compelling evidence that early bilinguals do not have intermediate, "compromise" phonetic categories, but rather display speech productions that are, by and large, no different from the input of the speech community to which they have been exposed. Taken together, the results support models of bilingual phonology that posit separation between phonological systems and are subject to dynamic phonetic interactions via language activation.

Supplementary Materials: The following are available online at https://www.mdpi.com/2226-471X/6/1/9/s1, All supplementary analyses as well as the data, code, and the experimental materials necessary to reproduce the analyses reported in this article are available at: https://osf.io/un45x/.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of the University of Arizona (Approval 12-0250-02).

Informed Consent Statement: All participants gave their informed consent for inclusion before they participated in this study.

Data Availability Statement: The complete dataset presented in this article is openly available on Open Science Framework and can be accessed at: https://osf.io/un45x/.

Acknowledgments: I express my gratitude to Miquel Simonet for sharing the dataset used in the analysis of coronal stops, as well as Kyle Jones for sharing data from his dissertation research. I am grateful for insightful comments from 3 anonymous reviewers that improved the quality of this work. All errors are mine alone.

Conflicts of Interest: The author declares no conflict of interest.

References

Amengual, Mark. 2012. Interlingual influence in bilingual speech: Cognate status effect in a continuum of bilingualism. *Bilingualism* 15: 517. [CrossRef]

Antoniou, Mark, Catherine T. Best, Michael D. Tyler, and Christian Kroos. 2010. Language context elicits native-like stop voicing in early bilinguals' productions in both L1 and L2. *Journal of Phonetics* 38: 640–53. [CrossRef] [PubMed]

Antoniou, Mark, Catherine T. Best, Michael D. Tyler, and Christian Kroos. 2011. Inter-language interference in VOT production by L2-dominant bilinguals: Asymmetries in phonetic code-switching. *Journal of Phonetics* 39: 558–70. [CrossRef] [PubMed]

Bohn, Ocke-Schwen. 2020. Cross-language phonetic relationships account for most, but not all L2 speech learning problems: The role of universal phonetic biases and generalized sensitivities. In *Approaches to the Study of Sound Structure and Speech: Interdisciplinary Work in Honour of Katarzyna Dziubalska-Kołaczyk*. London: Routledge, pp. 171–84.

Brysbaert, Marc. 2020. Power considerations in bilingualism research: Time to step up our game. In *Bilingualism: Language and Cognition*. Cambridge: Cambridge University Press, pp. 1–6. [CrossRef]

Bullock, Barbara. E., Alameida J. Toribio, Verónica González, and Amanda Dalola. 2006. Language dominance and performance outcomes in bilingual pronunciation. In *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference*. Somerville: Cascadilla Proceedings Project, pp. 9–16.

Bürkner, Paul-Christian. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80: 1–28. [CrossRef]

Caramazza, Alfonso, Grace H. Yeni-Komshian, Edgar B. Zurif, and Etorre Carbone. 1973. The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *The Journal of the Acoustical Society of America* 54: 421–28. [CrossRef]

Casillas, Joseph V., Yamile Díaz, and Miquel Simonet. 2015. Acoustics of Spanish and English coronal stops. Paper presented at 18th International Congress of Phonetic Sciences, Glasgow, UK, August 10–14; pp. 1–5.

Chang, Charles B., Yao Yao, Erin F. Haynes, and Russell Rhodes. 2011. Production of phonetic and phonological contrast by heritage speakers of Mandarin. *The Journal of the Acoustical Society of America* 129: 3964–80. [CrossRef]

Cho, Taehong, and Peter Ladefog. 1999. Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics* 27: 207–29. [CrossRef]

Chodroff, Eleanor, and Melissa Baese-Berk. 2019. Constraints on variability in the voice onset time of L2 English stop consonants. Paper presented at 19th International Congress of Phonetic Sciences, Melbourne, Australia, August 5–9; pp. 3348–3352.

Chodroff, E., and C. Wilson. 2017. Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics* 61: 30–47. [CrossRef]

Cohen, Jacob. 2013. *Statistical Power Analysis for the Behavioral Sciences*. London: Routledge.

Ellis, Paul D. 2010. *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results*. Cambridge: Cambridge University Press.

Escudero, Paola. 2005. Linguistic Perception and Second Language Acquisition. Ph.D. Thesis, Utrecht University, Utrecht, The Netherlands.

Flege, James. E. 1991. Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *The Journal of the Acoustical Society of America* 89: 395–411. [CrossRef]

Flege, James. E. 1995. Second language speech learning: Theory, findings, and problems. In *Speech Perception and Linguistic Experience Issues in Cross-Language Research*. Edited by W. Strange. Timonium: York Press, pp. 229–73.

Flege, James. E., and Wieke Eefting. 1987b. Production and perception of English stops by native Spanish speakers. *Journal of Phonetics* 15: 67–83. [CrossRef]

Flege, James. E., and Wieke Eefting. 1988. Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation. *The Journal of the Acoustical Society of America* 83: 729–40. [CrossRef] [PubMed]

Flege, James. E., and James Hillenbrand. 1984. Limits on phonetic accuracy in foreign language speech production. *The Journal of the Acoustical Society of America* 76: 708–21. [CrossRef]

Fowler, Carol. A., Valery Sramko, David J. Ostry, Sarah A. Rowland, and Pierre Hallé. 2008. Cross language phonetic influences on the speech of French–English bilinguals. *Journal of Phonetics* 36: 649–63. [CrossRef] [PubMed]
Gabriel, Christoph, Tanja Kupisch, and Jeanette Seoudy. 2016. VOT in French as a foreign language: A production and perception study with mono-and multilingual learners (German/Mandarin-Chinese). *SHS Web Conf.* 27: 1–14. [CrossRef]
Gelman, Andrew, Daniel Simpson, and Michael Betancourt. 2017. The prior can often only be understood in the context of the likelihood. *Entropy* 19: 555. [CrossRef]
Gertken, Libby M., Mark Amengual, and David Birdsong. 2014. *Assessing Language Dominance with the Bilingual Language Profile*. Edited by P. Leclercq, A. Edmonds and H. Hilton. Bristol: Multilingual Matters.
Gonzales, Kalim, and Andrew J. Lotto. 2013. A Bafri, un Pafri: Bilinguals' Pseudoword identifications support language-specific phonetic systems. *Psychological Science* 24: 2135–42. [CrossRef]
Gonzales, Kalim, Krista Byers-Heinlein, and Andrew J. Lotto. 2019. How bilinguals perceive speech depends on which language they think they're hearing. *Cognition* 182: 318–30. [CrossRef]
Grosjean. 2001. The Bilingual's Language Modes. In *One Mind, Two Languages: Bilingual Language Processing*. Edited by Janet Nicol. Oxford: Oxford University Press, pp. 1–22.
Hazan, Valerie L., and Georges Boulakia. 1993. Perception and production of a voicing contrast by French-English bilinguals. *Language and Speech* 36: 17–38. [CrossRef]
Holliday, Jeffrey J. 2015. A longitudinal study of the second language acquisition of a three-way stop contrast. *Journal of Phonetics* 50: 1–14. [CrossRef]
Jones, Kyle. S. 2020. A Multigenerational Investigation of Voice Onset Time in English-Hebrew Heritage Speakers. Ph.D. Thesis, The University of Arizona, Tucson, AZ, USA.
Kehoe, Margaret M., Conxita Lleó, and Martin Rakow. 2004. Voice onset time in bilingual German-Spanish children. *Bilingualism* 7: 71. [CrossRef]
Kiliç, Mehmet. 2018. Voice onset time as an indicator of interlanguage: Evidence from Turkish-accented English. *Journal of Language and Linguistic Studies* 14: 272–84.
Kilpatrick, Cynthia. D. 2004. Compromised VOT: Variation in a Bilingual Community. Ph.D. Thesis, University of Texas at El Paso, El Paso, TX, USA.
Kruschke, John K. 2018. Rejecting or accepting parameter values in Bayesian estimation. *Advances in Methods and Practices in Psychological Science* 1: 270–80. [CrossRef]
Kupisch, Tanja, and Conxita Lleó. 2017. Voice onset time in German-Italian simultaneous bilinguals: Evidence on cross-language influence and markedness. In *Romance-Germanic Bilingual Phonology*. Edited by M. S. Yavaş, M. Kehoe and W. Cardoso. Sheffield: Equinox, pp. 79–98.
Lein, Tatjana, Tanja Kupisch, and Joost van de Weijer. 2016. Voice onset time and global foreign accent in German–French simultaneous bilinguals during adulthood. *International Journal of Bilingualism* 20: 732–49. [CrossRef]
Lisker, Leigh, and Arthur. S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20: 384–422. [CrossRef]
Llama, Raquel, and Luz P. López-Morelos. 2016. VOT production by Spanish heritage speakers in a trilingual context. *International Journal of Multilingualism* 13: 444–58. [CrossRef]
López, V. G. 2012. Spanish and English word-initial voiceless stop production in code-switched vs. Monolingual structures. *Second Language Research* 28: 243–63. [CrossRef]
Lozano-Argüelles, Cristina, Laura Fernández Arroyo, Nicole Rodríguez, Ezequiel M. Durand López, Juan J. Garrido Pozú, Jennifer Markovits, Jessica Varela, Núria de Rocafiguera, and Joseph V. Casillas. 2020. Conceptually cued perceptual categorization in adult L2 learners. *Studies in Second Language Acquisition*, 1–16.
MacLeod, Andrea A., and Carol Stoel-Gammon. 2005. Are bilinguals different? What VOT tells us about simultaneous bilinguals. *Journal of Multilingual Communication Disorders* 3: 118–27. [CrossRef]
Magloire, Joël, and Kerry P. Green. 1999. A cross-language comparison of speaking rate effects on the production of voice onset time in English and Spanish. *Phonetica* 56: 158–85. [CrossRef]
Morgan, Dan W. 2011. Resetting VOT in a Bilingual Region. Ph.D. Thesis, The University of Texas at El Paso, El Paso, TX, USA.
Olson, Daniel J. 2013. Bilingual language switching and selection at the phonetic level: Asymmetrical transfer in VOT production. *Journal of Phonetics* 41: 407–20. [CrossRef]
Plonsky, Luke, and Frederick L. Oswald. 2014. How big is "big"? Interpreting effect sizes in L2 research. *Language Learning* 64: 878–912. [CrossRef]
R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, Available online: https://www.R-project.org/ (accessed on 10 October 2020).
Rao, Rajiv, and Rebecca Ronquest. 2015. The heritage Spanish phonetic/phonological system: Looking back and moving forward. *Studies in Hispanic and Lusophone Linguistics* 8: 403–14. [CrossRef]
Rohatgi, Ankit. 2020. WebPlotDigitizer: Version 4.3. Available online: https://automeris.io/WebPlotDigitizer (accessed on 10 October 2020).
Sakai, Mari. 2018. Moving towards a bilingual baseline in second language phonetic research. *Journal of Second Language Pronunciation* 4: 11–45. [CrossRef]

Schmidt, Anna M., and James E. Flege. 1996. Speaking rate effects on stops produced by Spanish and English monolinguals and Spanish/English bilinguals. *Phonetica* 53: 162–79. [CrossRef] [PubMed]

Simonet, M. 2014. Phonetic consequences of dynamic cross-linguistic interference in proficient bilinguals. *Journal of Phonetics* 43: 26–37. [CrossRef]

Stan Development Team. 2018. Stan Modeling Language Users Guide and Reference Manual (Version 2.18.0). Available online: http://mc-stan.org (accessed on 10 October 2020).

Stölten, Katrin, Niclas Abrahamsson, and Kenneth Hyltenstam. 2015. Effects of age and speaking rate on voice onset time: The production of voiceless stops by near-native L2 speakers. *Studies in Second Language Acquisition* 37: 71–100. [CrossRef]

Sundara, Megha, Linda Polka, and Shari Baum. 2006. Production of coronal stops by simultaneous bilingual adults. *Bilingualism* 9: 97–114. [CrossRef]

Van Leussen, Jan-Willem, and Paola Escudero. 2015. Learning to perceive and recognize a second language: The L2LP model revised. *Frontiers in Psychology* 6: 6–12. [CrossRef]

Walker, Abby. 2020. Voiced stops in the command performance of Southern US English. *The Journal of the Acoustical Society of America* 147: 606–15. [CrossRef]

Wan, Xiang, Wengian Wang, Jiming Liu, and Tiejun Tong. 2014. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Medical Research Methodology* 14: 135. [CrossRef]

Williams, Lee. 1977. The perception of stop consonant voicing by Spanish-English bilinguals. *Perception & Psychophysics* 21: 289–97.

Williams, Lee. 1980. Phonetic variation as a function of second-language learning. *Child Phonology* 2: 185–215.

Winter, B. 2011. Pseudoreplication in phonetic research. Paper presented at 17th International Congress of Phonetic Sciences, Hong Kong, China, August 17–21; pp. 2137–40.

Yazawa, Kakeru, James Whang, Mariko Kondo, and Paola Escudero. 2019. Language-dependent cue weighting: An investigation of perception modes in L2 learning. *Second Language Research* 36: 1–25. [CrossRef]

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
Tel. +41 61 683 77 34
Fax +41 61 302 89 18
www.mdpi.com

Languages Editorial Office
E-mail: languages@mdpi.com
www.mdpi.com/journal/languages

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland

Tel: +41 61 683 77 34
Fax: +41 61 302 89 18

www.mdpi.com

ISBN 978-3-0365-0967-9