languages

Article

The Link Between Perception and Production in the Laryngeal Processes of Multilingual Speakers

Zsuzsanna Bárkányi [1,2,*] and Zoltán G. Kiss [3]

[1] School of Languages & Applied Linguistics, Faculty of Wellbeing, Education & Language Studies, The Open University, Milton Keynes MK7 6AA, UK
[2] Hungarian Research Centre for Linguistics, 1068 Budapest, Hungary
[3] Faculty of Humanities, ELTE Eötvös Loránd University, 1088 Budapest, Hungary; gkiss.zoltan@btk.elte.hu
* Correspondence: zsuzsanna.barkanyi@open.ac.uk

Abstract: The present paper investigates the link between perception and production in the laryngeal phonology of multilingual speakers, focusing on non-contrastive segments and the dynamic aspect of these processes. Fourteen L1 Hungarian, L2 English, and L3 Spanish advanced learners took part in the experiments. The production experiments examined the aspiration of voiceless stops in word-initial position, regressive voicing assimilation, and pre-sonorant voicing; the latter two processes were analyzed both word-internally and across word boundaries. The perception experiments aimed to find out whether learners notice the phonetic outputs of these processes and regard them as linguistically relevant. Our results showed that perception and production are not aligned. Accurate production is dependent on accurate perception, but accurate perception is not necessarily transferred into production. In laryngeal postlexical processes, the native language seems to play the primary role even for highly competent learners, but markedness might be relevant too. The novel findings of this study are that phonetic category formation seems to be easier than the acquisition of dynamic allophonic alternations and that metaphonological awareness is correlated with perception but not with production.

Keywords: L3 phonology; multilingual phonology; speech perception; postlexical processes; regressive voicing assimilation; pre-sonorant voicing; VOT; Spanish; Hungarian; English

check for updates

Academic Editors: Romana Kopečková and Anna Balas

Received: 12 August 2024
Revised: 30 January 2025
Accepted: 31 January 2025
Published: 5 February 2025

Citation: Bárkányi, Z., & Kiss, Z. G. (2025). The Link Between Perception and Production in the Laryngeal Processes of Multilingual Speakers. *Languages*, *10*(2), 29. https://doi.org/10.3390/languages10020029

1. Introduction

Most studies on second/third language speech focus on the acquisition of contrasting sounds and investigate production. Less is known about the implementation of allophonic alternations that are (potentially) neutralizing. Research is also scarce on the relationship between speech perception and production for multilingual speakers (Wrembel et al., 2022), especially in relation to non-contrastive dynamic phonological processes. Therefore, the focus of the present study is to explore the link between these two domains in the laryngeal phonology of multilingual speakers. We examine the acquisition of a new phonetic category (aspiration) and the dynamic aspect of two postlexical assimilation processes. The paper also investigates individual speaker variation in these.

Wrembel et al. (2022) sum up four possible scenarios for the link between L2 speech perception and production. (1) L2 perception precedes (and surpasses) L2 production, that is, accurate production is only possible after accurate perceptual representations are formed; (2) perception and production co-evolve, especially in the early stages of acquisition; (3) accurate production precedes accurate perception, especially in the latter stages of learning when learners have more knowledge of the target language; and (4) there is no direct link between the two domains, especially when the L2 sounds are challenging from a motor-articulatory aspect. Since the participants of this study are competent speakers in their non-native languages and the sounds under scrutiny are not regarded to be challenging for articulatory execution (see below and in Section 4.1), we will consider the following possibilities: (1) the sound category/process is neither perceived or produced, i.e., it is not learned at all; (2) the sound/process is perceived but not produced (this would be scenario 1 for Wrembel et al. (2022)); (3) the sound/process is both perceived and produced, i.e., it is learned even if production might be not completely target-like, that is, an intermediate category might be formed; (4) the sound/process is not perceived but produced (scenario 3 above).

It has been widely discussed in L2 speech research why some sounds are harder to learn than others. Flege's Speech Learning Model (Flege, 1995) and its revised version (SLM-r) (Flege & Bohn, 2021), among others, assert that L2 speech learning is influenced by L1 phonetic system biases. According to this latter model, production and perception co-evolve in a bidirectional, though imperfect, relationship. The model treats cross-linguistic influence (CLI) as an equivalence classification, where L2 sounds are compared to L1 categories at the position-sensitive allophone level. Categories similar in L1 and L2 are harder to acquire as new categories, requiring learners to detect phonetic differences and break the L2-to-L1 perceptual link. Phonetic categories help access segment-sized units used during lexical access. This delinking can be hastened by expanding the L2 lexicon. While SLM-r focuses on individual sounds, other models, like L2LP (Escudero, 2005) and PAM-L2 (Best, 1994; Best & Tyler, 2007), focus on phonemic contrasts. L2LP, like SLM-r, posits that acoustic information predicts cross-linguistic categorization. It involves a perception grammar that maps acoustic signals to phonological representations. The model acknowledges a dissociation between perception and lexical encoding, i.e., linking phonological categories to lexical representations.

Representational approaches have often explained acquisitional difficulty in terms of markedness. Eckman (1985) argues that a structure which is absent from L1 and more marked than the L1 structure will be difficult to acquire, while a structure that is although absent from L1 but less marked than the L1 structure is easier to learn. For instance, it is easier to learn final devoicing for English speakers than final voicing contrast for German speakers. Young-Scholten (2004) on the other hand claims that it is not so much markedness but rather the domain of application of the phonological rule that matters. If an L2 process applies in a wider domain, a superset of the L1 phonological domain, it will be easier to acquire due to positive evidence. However, when the L2 domain is smaller (e.g., German word-final devoicing as opposed to the alveolar flapping rule in American English), it is more difficult to learn the alternation. In her study the allophonic alternation between the palatal and velar fricatives [ç]~[x] in German was easily acquired by her three L1 American English participants, while word-final devoicing was not learnt at all. We believe that the difference between the two allophonic rules might also arise because L1 laryngeal systems seem to be more difficult to overcome compared to acquiring alternations or differences based on place of articulation (e.g., Wrembel et al., 2020).

The realization of initial voiceless stops, specifically voice onset time (VOT) following the closure phase, has also received considerable attention in L2 and L3 speech as it is commonly associated with perceived foreign accent. In most studies dealing with VOT, learners had to acquire a new phonetic category, that is, they had to finetune the implementation of the existing voiceless–voiced contrast. The literature shows significant evidence of VOT transfer from the learners' L1 (e.g., Flege, 1987), which SLM explains by the equivalence classification, a mechanism that prevents the formation of new phonetic categories. Advanced learners also tend to produce hybrid in-between VOT values rather than target-like realizations (e.g., Flege, 1991). L3 studies also corroborate hybrid values and a combined influence from L1 and L2 (see Wrembel, 2015 for a detailed discussion), and even regressive transfer from L2/L3 (Sypiańska, 2013; Dmitrieva et al., 2020).

The languages involved in this study are L1 Hungarian, L2 English, and L3 Spanish. Hungarian is a true voice language (Beckman et al., 2013), where the voicing contrast of obstruents is based on negative voice onset time (VOT), or voice lead, in voiced stops vs. zero/short-lag VOT in voiceless stops. The language displays regressive voicing assimilation (RVA): adjacent obstruents must agree in their voicing feature, that is, voiced obstruents voice preceding voiceless obstruents (1a), and voiceless obstruents devoice preceding voiced obstruents (1b). The process is right-to-left iterative (1c). Hungarian has a symmetrical obstruent system with contrastive voiceless–voiced pairs at each place of articulation, including the word-final position (1d) (for details, see Siptár & Törkenczy, 2000).

(1)
a. /tb/ → [db]: e.g., *hát-ba* "back-ill"; *két#barát* "two friends"
b. /zt/ → [st]: e.g., *víz-től* "water-abl"; *víz#torony* "water tower"
c. /ɡdh/ → [kth]: e.g., *smaragd-hoz* "emerald-allat"
d. *méz* [z] "honey" ~ *mész* [s] "limestone"

English, just like Hungarian, displays a symmetrical laryngeal obstruent system, but unlike Hungarian and Spanish, English is an aspirating language (Lisker & Abramson, 1964), that is, the contrast of stops is based on aspiration rather than voicing. "Voiced" stops, or as they are generally referred to in the phonological literature, lenis stops (in initial position) are typically produced with zero or short-lag VOT, though negative VOT is also attested (e.g., Flege, 1982); thus, phonetically they are typically voiceless and unaspirated, while voiceless, or fortis, stops are produced prevocalically with a relatively long-lag VOT (i.e., aspirated). In contrast to true voice languages, in English, no systematic laryngeal spreading, i.e., RVA, is attested (Jansen, 2004; Szigetvári, 2020). Similarly to Hungarian, in English there is no pre-sonorant voicing either (PSV), that is, sonorant consonants do not voice preceding obstruents.

Spanish is also a true voice language, where stop phonemes can be either voiced or voiceless, although voiced stops are often realized as voiced approximants (unlike in Hungarian), and fricatives and affricates do not display a symmetrical system of laryngeal contrast. Even though Spanish has RVA, because of the phonotactic restrictions of the language, the segment undergoing assimilation is mostly /s/ (2a). Spanish presents a special case also because there is no alveolar voiced fricative phoneme in the language, and /s/-voicing only occurs in dialects where syllable-final /s/ rarely undergoes aspiration and deletion (Hualde, 2005). Importantly, unlike in Hungarian and English, /s/-voicing also occurs before sonorant consonants, including glides within the same word or across a word boundary, and although the process is highly variable, as a result, /s/ often becomes partially or fully voiced (2b). We summarize the relevant features of the laryngeal system of the three languages in Table 1.

(2)
a. /sd/ → [zð] *de**sd**e* "from"; /sb/ → [zβ] *coches#baratos* "cheap cars"
b. /sl/ → [zl] *isla* "island"; /sm/ → [zm] *las#minas* "the mines"

Table 1. Summary of the three laryngeal systems.

	Hungarian	**English**	**Spanish**
Type of laryngeal contrast	**True voice**	Aspirating	**True voice**
Laryngeal contrast within the obstruent inventory	**Symmetrical**	**Symmetrical**	**Symmetrical** (but limited to stops only)
RVA	**Yes**	No	**Yes**
PSV	**No**	**No**	Yes

In this paper, we focus on how different laryngeal phonological processes are produced and perceived by multilingual speakers. Specifically, we will take a closer look at the following areas:

1. Regressive voicing assimilation (between adjacent obstruents), which is present in participants' L1 and in most true voice languages, including their L3 Spanish.
2. Pre-sonorant voicing, which is not present in participants' L1 and is a typologically uncommon process. It involves extending RVA to sonorant triggers. Both 1 and 2 are dynamic phonological processes, and neither of them creates new segments for L1 Hungarian speakers.
3. Aspiration of voiceless stops in English, in which case a new phonetic category but no new phonological contrast has to be learned for L1 Hungarian speakers. Aspiration can be regarded as a word-level feature; thus, no dynamic alternation is learnt.

We posit that the segments that the participants of this study had to produce do not pose a motor-articulatory difficulty: RVA and PSV do not create new segments, and al-though voiceless stops require changing the timing of the onset of phonation, these segments are generally regarded as being fairly easy to articulate (Lock, 1972; Shuster & Cottrill, 2015). The following hypotheses are tested in the study:

Hypothesis 1. *All three processes under scrutiny are perceived by the highly competent advanced learners of this study. That is:*

a. *RVA in English is perceived as non-target-like.*
b. *PSV in Spanish is perceived as target-like, i.e., the lack of PSV in Spanish is perceived as non-target-like.*
c. *Aspiration of voiceless stops in English (long-lag VOT) is perceived as target-like, i.e., the lack of aspiration in English is perceived as non-target-like.*

Hypothesis 2. *RVA is likely to be blocked in English (since for aspirating languages progressive devoicing rather than RVA is the unmarked laryngeal scenario).*

Hypothesis 3. *PSV in Spanish is unlikely to be learned (since it is a typologically uncommon, marked process, and due to the lack of voicing contrast in Spanish fricatives, its realization is highly variable).*

Hypothesis 4. *Aspiration (long-lag VOT) in English is likely to be attested in the speech of these advanced learners (in accordance with previous studies that show advanced learners moving away from their native VOT patterns).*

Hypothesis 5. *Dynamic alternations are more likely to be produced within the word than across the word boundary (since, in the former case, word-level acquisition and encoding is sufficient).*

a. *RVA is more likely to be blocked within the word than across the word boundary.*
b. *PSV is more likely to be produced within the word than across the word boundary.*

2. Materials and Methods

2.1. Participants

Fourteen young adult subjects (5 male, 9 female) participated in the experiment, ranging in age from 19 to 25 years (average: 21.6). Each participant received a voucher of HUF 5000 for their involvement. All the subjects were students majoring in Spanish and English language and literature at Eötvös Loránd University, Budapest. They were native Hungarian speakers who started learning English and Spanish after the age of 11. Their proficiency in both languages was at least B2 according to the Common European Framework of Reference for Languages, as they all successfully passed both English and Spanish proficiency exams administered by the University as part of their studies. As they were all majoring in both languages, but did not study linguistics as a major or minor, they briefly learned about RVA, PSV, and aspiration during their studies. None had spent more than three months in an English-speaking or Spanish-speaking country, making them proficient sequential trilingual speakers who acquired their L2 English and L3 Spanish in a non-immersion context. Additionally, 10 subjects spoke another Romance language (3 spoke basic Portuguese, 3 spoke basic Italian, 2 spoke basic French, 1 person spoke basic Catalan and Portuguese, and 1 participant claimed to speak intermediate Italian). All considered themselves less proficient in the other Romance language, compared to English and Spanish. All participants claimed to speak the Northern–Central Peninsular variety of Spanish; 4 identified with American English, 6 with British English, and 4 with a mixed variety. None of the participants reported any speaking or hearing disorders.

2.2. Production Experiment: Materials, Measurements, and Statistical Analysis

The production experiment included two sessions: an English and a Spanish session. A reviewer pointed out our results would be more robust had we recorded Hungarian benchmark data from these participants. Unfortunately, due to time and resource constraints this was not possible, hence, for Hungarian, we have to rely on findings from previous research, a shortcoming that has to be kept in mind when interpreting the results. As language mode might have an effect on the acoustic realization of sounds (Amengual, 2021), and language mode was not a variable tested in the present research, the two sessions were kept separate. Half of the participants started with the English session and the other half started with the Spanish session. Since sessions were recorded on the same day, after the first session participants had a lunch break, and then they came back for the second session. Before each session they had to read a few sentences in Hungarian to adjust the microphone settings and to make sure that learners did not have any speech disorder. This might have resulted in a risk of L1 priming, a limitation that must be kept in mind when interpreting the data, but the sessions were long enough to counterbalance it. Sessions took place in the sound-attenuated booth of the Hungarian Research Centre for Linguistics. The production experiment overall involved a reading task, which means that the moderating role of orthography should be kept in mind. The participants had to read the test sentences and fillers from a monitor screen in a randomized order, which was generated by SpeechRecorder (Draxler & Jänsch, 2004). Each test sentence was read four times, and each participant was presented with a different randomized order in each round. The complete list of the test sentences for both languages can be found in the Appendix A.

The test material was used to investigate the following: RVA in English, PSV in Spanish, and VOT in English. The acoustic analysis of all the recorded sentences was carried out in Praat (v. 6.2.23, Boersma & Weenink, 2022). All the sentences were recorded in mono with a sampling rate of 44.1 kHz, which was resampled at 22.05 kHz and low-pass filtered with a cut-off at 11.025 kHz during the acoustic analysis. The recordings were segmented manually by the authors and a research assistant. We will detail the methodological peculiarities of each analysis in the following subsections.

2.2.1. RVA in English and PSV in Spanish

In the English RVA part of the production experiment, the target segment was /s/ within the word and word-finally, followed by a voiced stop /b/, /d/ or /ɡ/. Stimuli were embedded into 10–13-syllable-long neutral declarative carrier sentences. They occurred in the first half of the sentence, but were not sentence-initial. In the English session, there were 12 test sentences, each of which the 14 participants read out 4 times; 5 recordings had to be discarded due to errors, and, thus, we obtained a total of 667 recordings to analyze. The Spanish session contained 23 test sentences read out by the same 14 participants in 4 rounds; 5 sentences were discarded due to errors; thus, altogether 1283 recordings could be analyzed. As syllable count and stress were not variables included in this study, they were not controlled for. Although in Spanish and Spanish interlanguage voicing assimilations might be influenced by these factors (not in English to the best of our knowledge), our materials are adequate to test whether learners perceived and produced these processes. The test words included both cognates and non-cognates. On the selection of cognate words and their potential phonetic effect, see Bárkányi and G. Kiss (2024). The following measurements were carried out on the basis of the boundaries inserted:

1. Duration of the target consonant /s/.
2. Absolute length of the voiced interval.
3. Ratio of the voiced part compared to the total length of the consonant.

Although several other systematically occurring phonetic–acoustic correlates of voicing contrast are attested in the literature, such as voiced obstruents being generally shorter than their voiceless counterpart, and the preceding vowel being typically shorter before voiceless obstruents (see, e.g., House & Fairbanks, 1953; Chen, 1970; Kluender et al., 1988), in the current study only the proportion of voicing compared to the fricative interval was measured. The duration of the fricative was determined on the basis of the frication noise. Voicing was measured based on the visual inspection of the spectrograms and oscillograms, and a low-pass filter with a cut-off frequency of 500 Hz was used to securely determine the exact portion of the voicing oscillation during the frication noise.

In the Spanish PSV production experiment, the target segment was /s/ within the word and across the word boundary followed by a sonorant consonant /m/, /n/, and /l/. The presentation of stimuli, measurements, and statistical analysis were identical to those applied for RVA in English.

In Bárkányi and G. Kiss's (2023) production experiment, underlyingly voiced word-final /z/ in Hungarian before sonorants and before voiced obstruents was found to contain around 84% voicing on average, while final /s/ before voiced obstruents contained around 52% voicing. Their production experiment also investigated pre-sonorant word-final /s/, and found that it contained around 15% voicing on average.[1] Bárkányi and G. Kiss (2021, 2023) also showed that around 30% voicing in alveolar fricatives was the threshold for listeners to categorize them to be voiced across various phonetic environments, regardless of the duration of the preceding vowel. We will assume this voicing threshold in this paper, too.

2.2.2. VOT in English

This portion of the production experiment included words with initial /p/, /t/, and /k/. All words were at the very beginning of the test sentences. Each test word had two syllables; primary stress was always on the first syllable. A total of 8 test words were used (half of which were cognates) with 4 repetitions by 14 subjects; altogether, we had 448 data points for analysis.

The presence or absence of aspiration was detected by measuring the Voice Onset Time. Since Lisker and Abramson (1964), VOT has been used as one of the most established measurements of the laryngeal differences between (non-final) stop consonants, and it is defined as the timing relation between the moment of the release of the stop and the onset of glottal pulsing of the next vowel or sonorant consonant (Abramson & Whalen, 2017 provides a good overview of the theoretical and practical issues concerning VOT). This definition has become such a commonplace in the phonetic literature that many authors do not specifically indicate how exactly they measured VOT. However, this is problematic, as it is not straightforward where the analyst marks the beginning of VOT, i.e., the position of the release burst. Often there are multiple bursts or even releases without observable bursts. It is well known that posterior stops can have double or multiple bursts, and if all those are included in the VOT interval, this can result in significantly longer VOTs compared to anterior stops that often have fewer bursts or even no bursts at release. Some anterior articulations can even have multiple bursts, and if included in the VOT interval, this may result in longer VOT values and higher variation in the data. According to Keating et al. (1980), the release of the tongue from the palate is likely to induce the Bernoulli force making the organs reclose. In the case of posterior closure, higher intraoral pressure is more possible, making the reclosure more likely (also, the tongue dorsum has more inertia than the apex). The authors also suggest that at a later stage of the release, muscle force would have to play a role to maintain the high velocity airflow, acting against the Bernoulli force (1980: 5). We can take this as an indication that the final release burst (for anteriors, this is often the only burst) is the one that includes articulatory intention. For this reason, in our analysis, we measured VOT always from the last (or only) burst until the appearance of the first glottal pulse of the following vowel. This option also opened up the possibility for less variation in the VOT values (on the issue of multiple bursts in VOT measurement, see also Gráczi & Kohári, 2014). Using this principle, marking of the VOT interval was performed manually in Praat, based on the waveform and spectrogram, and a script then measured the duration of these intervals.

The inconsistency of marking release bursts may be one of the factors why the VOT results have been so varied in the literature, even for the same language. For example, in order to specify a boundary between non-aspirated (short-lag VOT) vs. aspirated (long-lag VOT) voiceless stops in Hungarian, we looked at the results of Lisker and Abramson (1964), Gósy (2000), Gósy and Ringen (2009), and a recent dissertation by Juhász (2024). While acknowledging that a straightforward comparison may not be possible due to differences in speech style (some works used isolated words lists, others spontaneous speech), position (initial vs. noninitial), the effect of the following vowel, etc., we can still see how varied the results for the same sounds are (Table 2).

Table 2. Mean VOT values (in ms) for Hungarian /p/, /t/, and /k/. LA 1964 = Lisker and Abramson (1964); G 2000 = Gósy (2000); GR 2009 = Gósy and Ringen (2009); J (2024) = Juhász (2024). The English VOT results in Lisker and Abramson (1964) have been added for comparison ("LA 1965 E").

Sound	LA 1964	G 2000	GR 2009	J 2024	Mean	LA 1964 E
/p/	2	24.6	9.7	33.25	17.39	58
/t/	16	23.3	16	32.45	21.94	70
/k/	29	50.1	37.6	67.65	46.10	80
Mean	15.6	32.7	21.1	44.45	**28.48**	69.3

The grand mean of the Hungarian VOTs turned out to be 28.48 ms. Despite the variation in the values, the means are actually rather close to what Keating (1984) established as the approximate cutoff points between the short-lag and long-lag VOT group: 20 ms for labials, 30 ms for apicals, and 40 ms for velars (1984: 298). In this paper, we have decided to mark the ceiling of short-lag VOT as Keating's (1984) upper threshold: 40 ms. English-type long-lag VOT (aspiration) is in the range of between 60 ms and 80 ms. VOT values between 40 ms and 60 ms will be regarded as intermediate.

2.2.3. Statistical Analysis of the Production Data

The statistical analysis (including the generation of the various plots and data tables) was carried out in *R* (v. 4.2.2, R Core Team, 2022) using various *tidyverse* packages (Wickham et al., 2019). Linear mixed-effects models were used to model the production data, using the packages *lm4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2020). The model function used the default settings (e.g., the Satterthwaite approximation was used to calculate the degrees of freedom for the *t*-distributions). Which exact predictor fixed effect was used in which model will be detailed in Section 3. The random-effect structure of the models contained the subject and item (i.e., the test words used in the experiments). Which exact model with what specific random-effect structure was used in which analysis will be detailed in the Results (Section 3). The best-fitting model was selected after carrying out model comparisons employing likelihood ratio tests (using maximum likelihood). A model was retained if the chi-square test was significant. The same procedure was used to test the utility of the random effects, except that in this case, restricted maximum likelihood was used in the likelihood ratio tests. Where more than two groups had to be compared, pairwise comparisons using Bonferroni adjustment to the *p*-values were carried out using the *emmeans* package (Lenth, 2023). The contributions of the random and fixed effects to explaining the variability in the data were measured by using marginal R-squared (abbreviated in the text as "R2m") and conditional R-squared ("R2c") (using the *r2* function of the performance package, Lüdecke et al., 2021).

2.3. Perception Experiments: Materials, Measurements, and Statistical Analysis

In order to explore to what extent the mentioned laryngeal processes are perceived by learners, a perception experiment was designed. While most L2 perception studies aim to determine if learners have acquired a specific contrast by using forced-choice tests (where participants have to decide whether they hear phoneme *A* or phoneme *B*), we found this method unsuitable for our study. This is partly because listeners might compensate for assimilations (cf. Bárkányi & G. Kiss, 2021, 2023), and partly because they might have a bias against segments not present in their phoneme inventory or be influenced by the orthographic form of the test words. Since speakers often remain unaware of postlexical processes even in their L1, we wanted to allow learners the option of responding "I don't know" or indicating they cannot perceive the processes under scrutiny at all or as linguistically relevant features.

To address this, we chose a more holistic approach. A short story, approximately one minute long, was recorded in Spanish and in English by two phonetically trained bilingual female speakers with native-like proficiency in both languages (Hungarian–English and Hungarian–Spanish, respectively; the texts can be found in the Appendix A). The same story was recorded again twice in English and once in Spanish: the first one with RVA applied in the English text as it would be in Hungarian (all the relevant tokens contained at least 30% of voicing); the second one with no aspiration, that is, short-lag (<40 ms) VOT. The Spanish text was re-recorded without pre-sonorant voicing to mirror the L1 laryngeal patterns. Participants listened to each of these slightly different recordings (three in English, two in Spanish) three times in a random order, resulting in a total of nine and six recordings, respectively, per language.

The experiment was conducted using Praat MFC with the same participants from the production experiments, at the end of the production sessions. While listening to the texts, the monitor screen remained blank. Texts were separated by 1.5 s of silence and a 5 s bell. After each text, instructions appeared on the screen, and participants had to rate how native-like the speaker sounded on a scale from 1 to 5 (1 being not at all native-like and 5 being completely native-like).

The perception experiment fitted cumulative link mixed models to the data using the *ordinal* package (Christensen, 2022). The outcome variable was the rating of nativeness by the participants, which was on an ordinal scale (1, 2, 3, 4, 5). The fixed predictor variable was the *stimulus* they listened to (with three levels for each language, English: native, non-native without English aspiration, non-native with Hungarian RVA applied; Spanish: native, non-native without sonorant voicing). The best-fitting model was defined for each language separately as: *clmm(rating ~ stimulus + (1+stimulus | subject), Hess = TRUE, link = "probit")*. The link function used was probit, as it is considered to be more suitable than the logit link function in models that contain random effects (Hahn & Soyer, 2005). The best-fitting model was selected using the same principles and procedure as in the case of the linear mixed-effects models used in the analysis of the production data.

2.4. Metaphonological Awareness

Metalinguistic awareness has long been recognised as a crucial element in the acquisition of foreign languages. Scholars often distinguish between intuitive and conscious language awareness or knowledge (e.g., James, 1999; Bialystok & Ryan, 1985). In order to better understand individual learner characteristics in the acquisition of non-contrastive segments and processes, a short metaphonological awareness (MPhA) test was administered to the participants of the study after they completed the production and perception experiments (Appendix A). The test was in Hungarian and focused on sound awareness (where letter-to-sound correspondence was not straightforward) (questions 1 and 3: worth 8 points); understanding the notion of syllable (question 2: 3 points); phonetic similarity between English, Spanish, and Hungarian sounds (questions 4–6(7): 3 points); phonotactic awareness (question 7: 1 point); and knowledge about lexical stress in the three languages (question 8: 1 point). In total, participants could gain 16 points during this test.

3. Results

3.1. Overall Production Results

3.1.1. RVA in English

We begin with the RVA results in the English words. As Figure 1 shows, there was a very similar amount of voicing in /s/ before voiced stops both word-internally and word-finally: in both environments, /s/ contained around 52% of voicing on average (Table 3 contains the relevant descriptive statistics). This amount of average voicing is actually very close to what Bárkányi and G. Kiss (2023) found for Hungarian word-final /s/ before a voiced stop: the overall mean there was around 52%, too, as we mentioned in Section 2.2.1. If we assume that around 30% of voicing in /s/ is sufficient to induce a voicing response perceptually (see Section 2.2.1), we can conclude that /s/ was mostly voiced in this data. This also indicates that, on average, /s/ in the English words was voiced as in the L1 of our participants. The distribution of voicing was also remarkably similar in the two environments: while around 63% of the observations were voiced at or above the 30% threshold (almost 30% of the observations were fully voiced), around 37% of the scores contained less than 30% voicing. This bimodal distribution indicates that certain speakers produced the English pattern, in which there is no RVA, while others did not. We will come back to the individual data in Section 3.4. The only minor difference between the two environments was in the ranges of the data: word-final /s/ always contained at least 10% voicing (i.e., there were no completely unvoiced tokens). The best-fitting linear mixed-effects model for the English RVA data was the one with varying by-subject intercepts and slopes and varying by-item (*word*) intercepts, with the following model structure: *lmer(voicing ~ 1 + environment + (1+environment | subject) + (1 | word))*. According to this model, the fixed effect (*environment*) is not statistically significant and had a low effect size, indicating no real and substantial difference between the two groups. The fixed effect alone, therefore, does not explain most of the total variance; the random effects (subject- and item-specific variability), however, play a sizable role in explaining it, as shown by R2c ($b = 1.22$, $t(15.2) = 0.417$, $p = 0.683$; R2m < 0.01, R2c = 0.52).

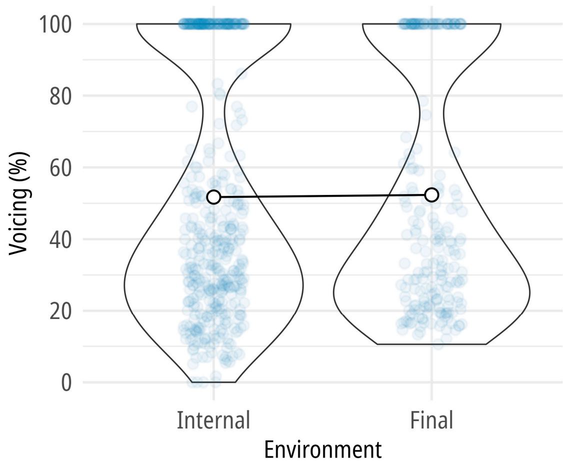

Figure 1. Voicing of /s/ before voiced stops in the English words. The white circles inside the violin plots represent the means.

Table 3. Descriptive statistics for the voicing of /s/ before voiced stops in the English words. B30 = number of observations below 30% voicing; A30 = number of observations above or equal to 30% voicing; 0 = number of observations with 0% voicing; 100 = number of observations with 100% voicing.

Environment	N	Mean	SD	Median	Min	Max	B30	A30	0	100
Internal	448	51.71	34.19	39.80	0.00	100	165 (36.83%)	283 (63.17%)	4 (0.89%)	131 (29.24%)
Final	219	52.32	33.41	39.62	10.55	100	84 (38.36%)	135 (61.64%)	0 (0%)	65 (29.68%)

3.1.2. PSV in Spanish

The Spanish pre-sonorant data show that mean voicing of /s/ was below 30% on average both word-internally and finally (see Figure 2 and Table 4). Almost 94% of the internal data contained less than 30% voicing, and around 81% of the word-final tokens had voicing below 30%. There was no word-final /s/ that was fully voiced, and only eight tokens were 100% voiced word-internally, which can be clearly regarded as outliers in the overall distribution. A few tokens were completely voiceless, too, in both positions (7% internally, 5% finally). The means found in this experiment are very similar to what Bárkányi and G. Kiss (2023) reported for Hungarian word-final pre-sonorant /s/, which contained around 15% voicing on average. The model that turned out to be the best fitted to the data contained by-subject intercepts and slopes, with the following model structure: *lmer(voicing ~ 1 + environment + (1+environment | subject))*. According to this model, the fixed effect (*environment*) is statistically significant; however, it alone does not explain a large portion of the total variance (only 6%); the random effect (subject-specific variability) plays a larger role in explaining the outcome variance (b = 7.066, t(14.004) = 3.825, p = 0.002; R2m = 0.06, R2c = 0.19). Again, it needs to be pointed out that despite the meaningful differences, both group means were below the 30% voicing threshold, rendering the /s/-realizations predominantly voiceless.

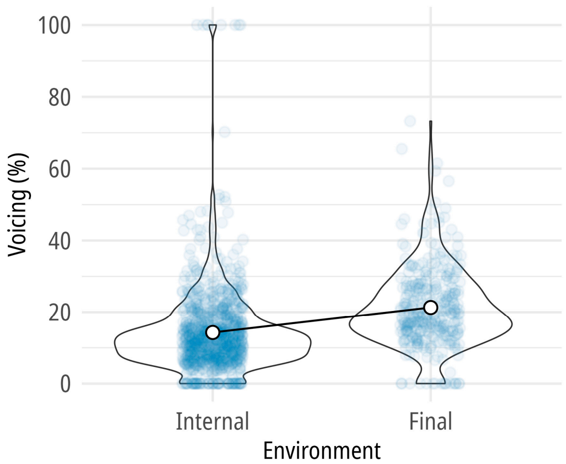

Figure 2. Voicing of /s/ before sonorants in the Spanish words. The white circles inside the violin plots represent the means.

Table 4. Descriptive statistics for the voicing of /s/ before sonorants in the Spanish words. B30 = number of observations below 30% voicing; A30 = number of observations above or equal to 30% voicing; 0 = number of observations with 0% voicing; 100 = number of observations with 100% voicing.

Environment	N	Mean	SD	Median	Min	Max	B30	A30	0	100
Internal	947	14.26	12.10	12.17	0	100.00	889 (93.88%)	58 (6.12%)	68 (7.18%)	8(0.84%)
Final	336	21.31	11.31	19.49	0	73.22	273 (81.25%)	63 (18.75%)	17 (5.06%)	0(0%)

3.1.3. VOT in English

The Voice Onset Time results are shown in Figure 3 (descriptive statistics are in Table 5). The first violin plot shows all the observations for all the three stops together, while the other three show the VOT scores for each stop separately. We can observe a rather stretched distribution of the data overall. The mean (37.45 ms) is at the 40 ms upper threshold we set for the short-lag VOT (unaspirated) category, but it is way below the 60 ms lower threshold of the long-lag VOT (aspirated) category. This indicates that a large chunk of the scores fell between these two categories. Indeed, around 26% of the data were between 40 and 60 ms (59% were below 40 ms, and around 15% were above 60 ms). Due to space restrictions and due to the fact that our focus is an overall quantification of whether our participants aspirated the stops or not, we do not detail the place of articulation differences in the VOT values; we just note here that they are in accordance with what has been widely observed in the literature: labial /p/ has the smallest mean VOT, velar /k/ has the highest, and alveolar /t/ occupies a place between the two.[2] Compared with the means we showed in Table 2, these means are about 1.5 times higher for /p/ and /t/, which reflects the relatively high number of intermediate values for these two sounds, too. The mean for /k/ was 0.95 times lower in our data than the mean of the literature reviewed, which may well be due to the inclusion of multiple bursts in the VOT in velars in the literature, discussed in Section 2.2.2.

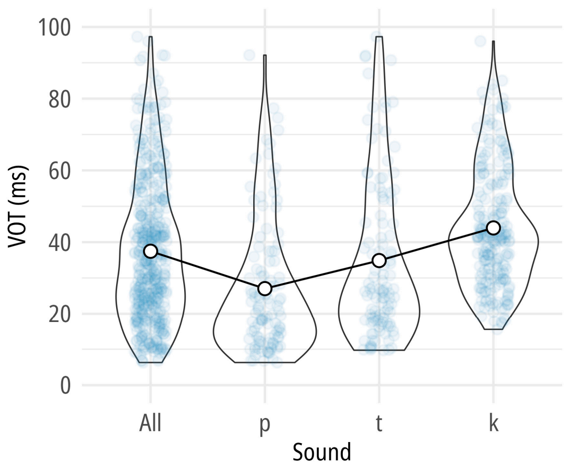

Figure 3. Voice onset time of word-initial /p/, /t/, and /k/ in the English words. The white circles inside the violin plots represent the means.

Table 5. Descriptive statistics for the VOT of /p/, /t/, and /k/ in the English words. B40 = number of observations whose VOT equals or is below 40 ms; IM = number of observations with an intermediate VOT between 40 and 60 ms; A60 = number of observations with a VOT above 60 ms.

Sound	N	Mean	SD	Median	Min	Max	B40	IM	A60
All	448	37.45	19.89	34.71	6.28	97.28	264 (58.93%)	116 (25.89%)	68 (15.18%)
p	112	27.03	18.52	21.31	6.28	92.14	90 (80.36%)	13 (11.61%)	9 (8.04%)
t	112	34.85	22.28	26.75	9.71	97.28	78 (69.64%)	19 (16.96%)	15 (13.39%)
k	224	43.95	16.57	41.57	15.54	96.03	96 (42.86%)	84 (37.5%)	44 (19.64%)

3.2. Overall Perception Results

We begin the perception results with the English data. Panel A in Figure 4 displays the ratings of the participants of the two texts. "Tom" was the native-like recording, "TomAsp" was the one where voiceless stops were pronounced with Hungarian-like VOTs (unaspirated), while "TomVoi" was the one where Hungarian-like RVA was applied (1 = not at all native-like, 5 = completely native-like). As we can see, the ratings were lower for the non-native-like recordings; for example, no participant ranked it with the highest score 5, and while there was barely any rating of 1 for the native recording, the non-native ones received ratings of 1 and 2.

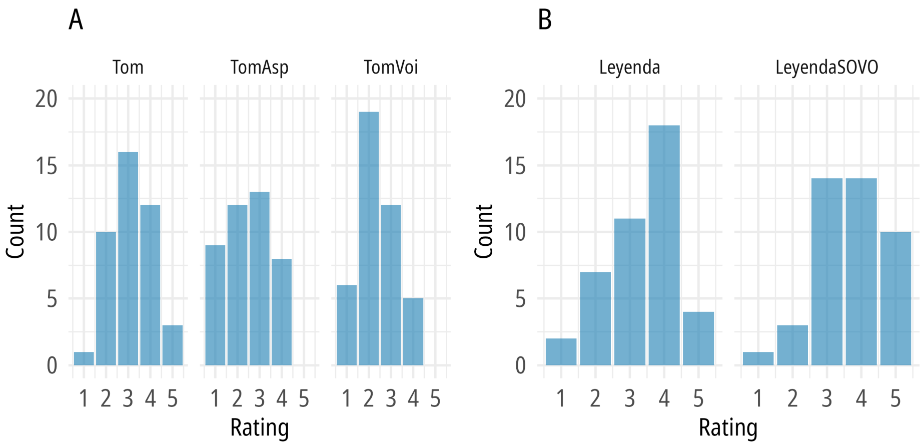

Figure 4. Rating of the English texts (**A**) and rating of the Spanish texts (**B**).

The best-fitting cumulative link mixed-effects model was the one which included *stimulus* as the fixed effect and by-subject varying intercepts (but not slopes); model structure: *clmm(rating ~ stimulus + (1 + stimulus | subject), Hess = TRUE, link = "probit")*. According to this model, compared to the native recording, the non-native recordings significantly decreased the ratings, i.e., lower ratings were more likely (TomAsp: $b = -1.02$, SE = 0.25, $z = -4.054$, $p < 0.001$; TomAsp: $b = -1.15$, SE = 0.26, $z = -4.475$, $p < 0.001$; R2m = 0.13, R2c = 0.53). These results indicate that the participants reliably differentiated between the native recording and each of the non-native ones, and that they rated the non-native ones (with either RVA applied or aspiration not applied) much lower. The R-squared values indicate that the fixed effect (*stimulus*) has a moderate effect size, while subject-related random effects contribute relatively more to the variance. Pairwise comparisons revealed no statistically significant difference between the ratings of the two non-native texts.

The ratings of the Spanish recordings can be seen in panel B of Figure 4. "Leyenda" was the native-like recording, "LeyendaSOVO" was the one in which no sonorant voicing was applied in the relevant phonological environments (hence, the latter was the non-native-like recording). The same model structure was used here as in the English recordings: the best-fitting model was the one which included by-subject varying intercepts (but not slopes), in addition to *stimulus* as the fixed effect. According to this model, while the non-native recordings increased the ratings, these increases were not statistically significant ("LeyendaSOVO": $b = 0.46$, SE = 0.24, $z = 1.9$, $p = 0.0564$).

3.3. Follow-Up Perception Experiment

The production and perception data have shown that the multilingual speakers of our experiment did not perceive and did not produce PSV in Spanish. This is somewhat surprising as they are advanced learners, highly competent speakers of Spanish. In their L1 Hungarian voicing is a distinctive feature before sonorant consonants (e.g., *ré**szn**ek* [sn] "part.DAT" vs. *ré**zn**ek* [zn] "copper.DAT"). However, it is precisely for this reason that the Hungarian voicing of /z/ in this position is not optional, unlike in Spanish.

Thus, to further explore whether L1 Hungarian–L2 English learners perceive PSV in Spanish, an AXB forced-choice follow-up experiment was designed. The stimuli consisted of six nonce word pairs of the form V*s*SonV. The consonant before the sonorant was the alveolar fricative /s/ either completely voiceless or partially voiced (25–35% voicing). One stimulus consisted of three words separated by a one-second silence, and participants had to decide whether the second word was more similar to the first one or to the third one. The order of the words within a stimulus, that is, AAB, ABB, BBA, BAA (e.g., usmo-usmo-u[z]mo, usmo-u[z]mo-u[z]mo, etc.), as well as the stimuli, were randomized across trials in PraatMFC. This task did not require phonological processing, but it did require noticing the phonetic difference between the examined segments. As goodness in this experiment only focused on accuracy, there was no response time limit, but participants could listen to a stimulus at most twice. After they responded, they heard a bell sound and then the next stimulus.[3]

To ensure naturalness and avoid artificial concatenation, stimuli were recorded by a phonetically trained female native speaker of Northen–Central Peninsular Spanish. Several items of the same stimulus were recorded, and where the manipulation sounded natural and also satisfied the measured parameters (voicing ratio), those were chosen by the authors and validated by an independent native speaker. Thirteen participants were available for this experiment. The rate of successful recognition was 96.9% overall. Pre-sonorant voicing was correctly recognised 308 times out of 312 stimuli (success rate % = 98.7). This means that PSV was reliably perceived by the participants.

3.4. Results at the Individual Level

We have seen in the production data that subject-dependent variability was a major factor. In this section we discuss to what extent individual cases support the group trends reported so far. In order to explore individual patterns in the relationship between perception and production, we decided to make categorical decisions, that is, we considered that a participant either perceived a process under scrutiny or they did not, and either realized a process in their speech or they did not. We are not claiming that the mastery of these processes in either domain is binary like this; nonetheless, we think this is a helpful first exploration to understand the link between perception and production in non-contrastive phonological processes.

The cut-off value to determine whether a learner blocked RVA in English was 30% of voicing during the fricative interval (see Section 2.2.1). Figure 5 shows the proportion of voicing in /s/ before voiced stops in word-internal and word-final (sandhi) positions by speakers.

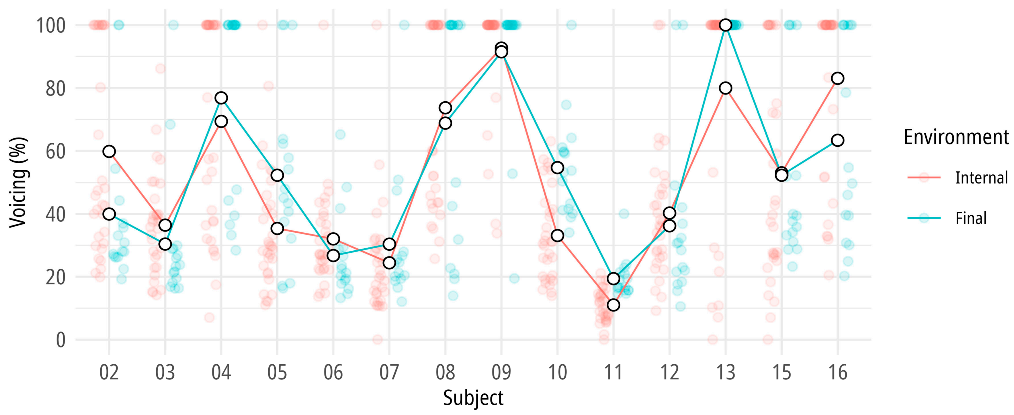

Figure 5. Voicing of /s/ before voiced stops in the English words by subjects. The white circles represent the means.

We can see that participant S11 is well below the 30% cut-off value, that S06 and S07 averaged across the word-internal and the sandhi context are below 30%, and that S03 is a borderline case. However, it turns out that for this speaker one word (*baseball*) increased the averages; without this word, their averages were also below 30%. We believe that these four speakers made an intended effort to block RVA in their English speech. It is worth mentioning that although these averages give a more nuanced picture, they still conceal the high intra-speaker variability observed in our data, which is shown by the presence of the individual data points in Figure 5 and below.

We used a cut-off value of 40 ms VOT for aspiration, as explained in Section 2.2.2. Values below this were considered Hungarian-like; above it, we assumed an intended articulatory gesture to achieve more target-like aspiration (Figure 6). The results show that six learners moved away from L1 short-lag VOT, but only two (S05 and S11) achieved a target-like aspiration.

Applying the same 30% of voicing as a cut-off value for PSV in Spanish, the results show that there were no participants who consistently produced pre-sonorant /s/ voicing in Spanish (Figure 7), although speakers S05 and S13 were close to this value in the sandhi context (29.91% and 30.17%).

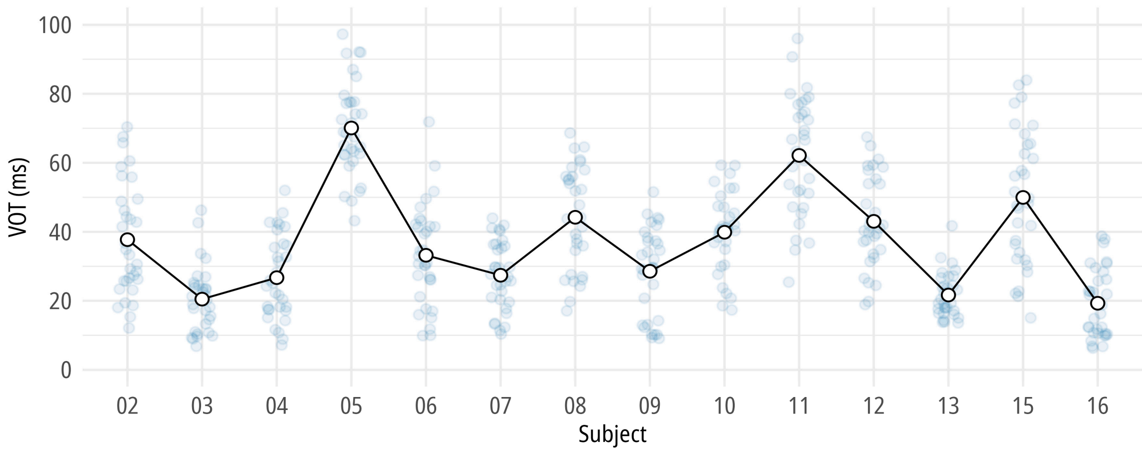

Figure 6. Voice Onset Time of word-initial /p/, /t/, and /k/ in the English words by subject. The white circles represent the means.

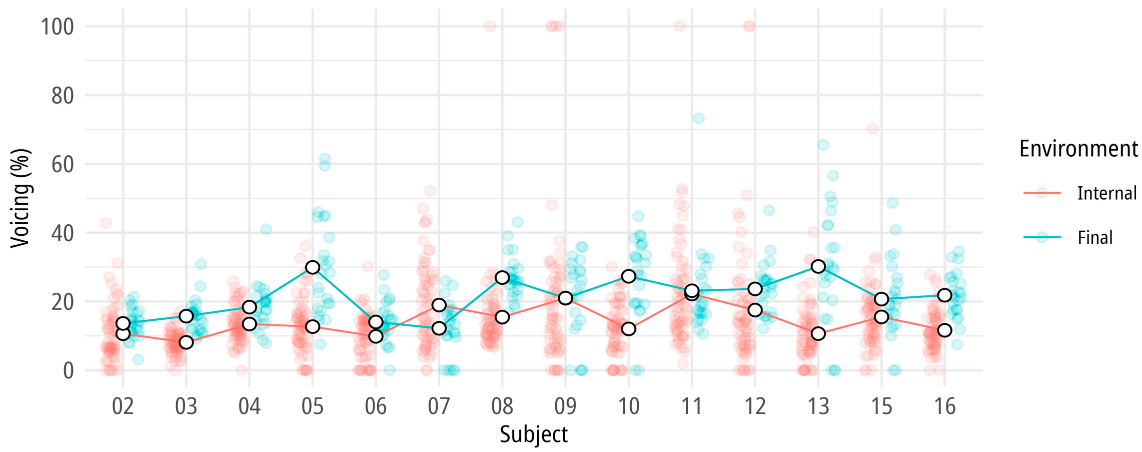

Figure 7. Voicing of /s/ before sonorants in the Spanish words. The white circles represent the means.

Decisions regarding the perception of the three laryngeal processes were made in the following way. If a participant rated the manipulated texts lower than the non-manipulated one, we assumed that they heard the incorrect application of RVA, or lack of aspiration in English, and the lack of pre-sonorant voicing in Spanish. If they considered the manipulated texts as good or better than the non-manipulated ones (this latter only occurred in relation to the Spanish texts), we deemed that they did not perceive the phonological processes under scrutiny. Based on these principles, the following individual patterns were attested (Table 6).

Based on Table 6, the following overall perception–production groups can be distinguished (Table 7). They reflect the perception–production link possibilities, discussed in Section 1. We can see that the largest group was possibility 1 when the sound/process is not learned at all, followed by the situation when the sound/process is perceived but not transferred to production. The group in which the sound/process was perceived as well as produced was smaller, and there was no participant who produced but did not perceive the sound/process under scrutiny.

Table 6. Individual patterns in perception and production (No = the process is not perceived/produced; Y = the process is perceived/produced; MPhA = metaphonological awareness).

	RVA in English		Aspiration in English		PSV in Spanish		MPhA Score
	perc.	prod. (block)	perc.	prod.	perc.	prod.	
S02	No	No	No	No	No	No	6
S03	Y	Y	No	No	No	No	6
S04	Y	No	Y	No	No	No	8
S05	No	No	Y	Y	No	No	5
S06	Y	Y	Y	No	No	No	10
S07	Y	Y	Y	No	No	No	8
S08	Y	No	Y	Y	No	No	8
S09	Y	No	Y	No	No	No	9
S10	Y	No	Y	Y	No	No	8
S11	Y	Y	Y	Y	Y	No	12
S12	Y	No	Y	Y	Y	No	14
S13	Y	No	No	No	No	No	8
S15	Y	No	Y	Y	No	No	5
S16	Y	No	Y	No	No	No	10

Table 7. Summary of the perception–production groups by process.

Pattern	RVA	VOT	PSV	Total
1. No perception–no production	2	3	12	17
2. Perception–no production	8	5	2	15
3. Perception–production	4	6	0	10
4. No perception–production	0	0	0	0

3.5. Metaphonological Awareness Results

With the aim of exploring the connection between metaphonological awareness (MPhA) and the acquisition of non-contrastive processes, we quantified the perception and production patterns. If a learner either perceived or produced a process, a score of 1 was assigned; if they both perceived and produced it, they received 2 points. This means that the score for perception and production for the four examined processes was 8. Then we measured the correlation between metaphonological awareness and the production and the perception scores, as well as both scores together (total) using Pearson's correlation coefficient r. We also fitted simple linear regression models to the data of each of the three domains, investigating the effect of MPhA on perception, production and both together. As the correlation coefficients and the model summaries indicate (Table 8), metaphonological awareness had a substantial effect on perception (and, overall, on perception and production), but not on production. Thus, all three models predict that for each increase in metaphonological awareness score by one, all three domain scores increase, but it is the perception score that increases the most (by 0.752), and this increase is statistically significant, too.

Table 8. Simple linear regression model results for metaphonological awareness and the three domains. The table also includes Pearson's *r* for the correlation between the variables.

Domain	r	Term	b	SE	df	t	p	R^2
Perception	0.752	MPhA	0.236	0.060	12	3.954	0.0019	0.566
Production	0.302	MPhA	0.095	0.086	12	1.099	0.2932	0.091
Total	0.601	MPhA	0.288	0.110	12	2.605	0.0230	0.361

4. Discussion

The main aim of this study was to explore the link between perception and production in the laryngeal phonology of multilingual speakers. We focused on the dynamic aspect of these processes (where relevant) and on individual patterns. We repeat the hypotheses here, along with the decision as to whether they can be rejected or retained based on the results; each will be detailed in the subsections below.

Hypothesis 1. *All three processes under scrutiny are perceived by the highly competent advanced learners of this study.*
The results **do not support** the hypothesis, because while RVA and the lack of aspiration were mostly perceived, PSV as a phonological process remained unnoticed.

Hypothesis 2. *RVA is likely to be blocked in English.*
The results **do not support** the hypothesis, because learners produced highly voiced /s/ tokens both within the word and across the word boundary.

Hypothesis 3. *PSV in Spanish is unlikely to be learned.*
The results **support** the hypothesis, as PSV was neither perceived nor produced by the participants of this study.

Hypothesis 4. *Aspiration (long-lag VOT) in English is likely to be attested in the speech of these advanced learners.*
The results **do not support** the hypothesis because less than half of the participants (six learners) moved away from L1 aspiration patterns in their productions, but only two approximated native-like VOT durations.

Hypothesis 5. *Dynamic alternations are more likely to be produced within the word than across the word boundary.*
The results are inconclusive because RVA tended to be blocked within the word, but less across the word boundary, while PSV was applied to a significantly greater degree across the word boundary than within the word.

4.1. Voicing Assimilation

It has been shown that the three laryngeal processes under scrutiny do not behave in the same way; thus, **H1** (*all three laryngeal processes are perceived by the highly competent advanced learners*) is not supported. RVA and aspiration were mostly perceived, while PSV remained unnoticed. The difference between the perceptions of the Hungarian-like application of RVA in English and the lack of PSV in Spanish is somewhat surprising, since both are postlexical assimilation processes that do not create new segments for L1 Hungarian learners. There might be several factors behind this. One is that /s/ and /z/ are contrastive in pre-sonorant position in Hungarian, rendering several minimal or quasi-minimal pairs (see Section 3.3), so learners might want to keep the two categories safely apart in this context; however, they encoded all Spanish words containing an alveolar fricative with a voiceless /s/. This might be corroborated by the production data (Figure 2), as pre-sonorant /s/ in Spanish was more voiced in the sandhi context than within the word, although the two group means were below the 30% voicing boundary (see Section 4.2). On the other hand, English words are encoded as containing either /s/ or /z/, and although in their productions, learners did not seem to be able to override their L1 laryngeal system (namely RVA), they did notice it as non-target-like in the speech of others. Another relevant factor, widely discussed in the literature, is the variable and gradient nature of PSV in Spanish (e.g., Bárkányi, 2014, 2018), which might obstruct noticing the process as linguistically relevant even though all the participants identified with Northern–Central Peninsular Spanish. Markedness might also have a role to play. RVA triggered by voiced obstruents is a common, unmarked process in true voice languages, such as Hungarian and Spanish, but not in aspirating languages, like English (Jansen, 2004). We might assume that a universally unmarked phonological process is easier to learn than a marked process (Eckman, 1985)—even if in this case "learning" is meant to be limited to the perceptual domain only. Further studies will have to explore whether speakers of languages which display /s/–/z/ contrast before sonorants within the word but neutralize it across word boundaries (like Slovak) perceive both processes in the same way. Additionally, research should explore whether markedness affects perception and production in the same way.

The occurrence of RVA in the participants' English interlanguage is interesting, and it means that **H2** (*RVA is likely to be blocked in English*) is not supported. It indicates that learners mostly perceive the non-target-like application of the process but are unable to prevent its use in their own speech. As the phonetic segments involved ([s] and [z]) are attested in participants' L1, motor-articulatory difficulty as a possible reason for this can be ruled out. A reviewer suggested that the bimodal distribution shown in Section 3.1 might not only indicate inter-speaker variation, but also optionality in learners' grammar. This might be the case for Speaker S15 and, potentially, for S04; however, the exploration of the interplay between individual patterns and lexical factors is beyond the scope of the present paper and will be worth investigating in the future.

Our results add to the limited body of evidence that L1 voicing processes are difficult to overcome (Bárkányi & G. Kiss, 2024; Wrembel et al., 2020). Escalante (2016), in a study with only six L1 English learners of Spanish, found that while the voicing of intervocalic /s/ (e.g., *presidente* "president") decreases with proficiency, no such development is observed in voicing assimilation (the study did not differentiate between RVA and PSV). Skarnitzl and Šturm (2017), in a study with Czech and Slovak L1 learners of English, also found that L1 RVA transfer is present in their English interlanguage, and that Slovak speakers also apply PSV while Czechs do not in accordance with their L1 voicing systems. The study also found that the tendency to assimilate in English correlates with the perception of foreign-accented speech. Further investigation should shed light on whether non-target-like voicing assimilation also correlates with intelligibility and whether it is worth explicitly teaching these phonological processes. These results also support the assertion that there is no direct correlation between perception and production (as described by Liu & Lin, 2021). Wrembel et al. (2022) reported perception and production not being aligned in the L3 of multilingual speakers. Our data have brought evidence in support of this, too; however, the authors found that the same learners showed aligned perceptual and productive performance in their L2 in which they were competent speakers (they were beginners in their L3). Since our participants are highly proficient multilingual speakers, it is unlikely that they will suppress RVA in their English interlanguage at a later stage and notice and apply PSV in Spanish.

4.2. Dynamic Alternations

Dynamic alternations are understudied in L2 (and L3) speech research. The two postlexical processes (RVA and PSV) showed a varied picture in our participants' speech. RVA overall was blocked slightly more successfully within the word than across the word boundary—where no fully voiceless realizations were attested—but the difference was not statistically significant. PSV, on the other hand, was significantly more attested in the sandhi context than within the word, but in both contexts, voicing was well below the 30% cut-off value. Thus, results concerning **H5** (*dynamic alternations are more likely to be produced within the word than across the word boundary*) are inconclusive.

A reason behind why there were slightly fewer cases of RVA in English within the word than in sandhi might be that both English and Hungarian orthography reinforce keeping voiced and voiceless categories apart. Thus, at word level, the voiceless segment is more likely to be encoded and lexicalized as such, potentially blocking RVA within the word. However, blocking RVA across the word boundary requires more than word-level acquisition; thus, voicing induced by RVA in sandhi is an L1 transfer, and is more difficult to unlearn. We hypothesize that word-level phonological knowledge might be more similar to declarative knowledge, while the mastery of dynamic processes might be more similar to procedural knowledge and, as such, needs more practice to become automatic.

Pre-sonorant voicing data seemingly contradict this since there were more voiced word-final /s/ realizations than within the word, which suggests that participants, even if to a limited degree, should tend to achieve the dynamic aspect of the process more successfully. We do not think, however, that this is the case. Our participants did not perceive PSV and, consequently, they did not learn to produce it at all; thus, **H3** (*PSV in Spanish is unlikely to be learned*) is borne out, we believe, for the reasons described above. The attested voicing surplus across the word boundary is probably the result of phonetic coarticulation, i.e., modal voicing rather than an intended articulatory gesture. Schmidt and Willis (2011), for instance, in Mexican Spanish, found considerable voicing even in an expected voiceless context: 0–34 ms in the intervocalic position and up to 37 ms preceding a voiceless consonant. In our data, voicing duration in /s/ before sonorants was only 12.16 ms on average within the word and 14.02 ms across the word boundary. This amount of voicing does not seem to be consistent with planned, non-coarticulatory voicing, as in RVA. In the Spanish interlanguage of L1 Hungarian learners, there is only one alveolar sibilant, which is voiceless, and they seem to intend to realize it as such in their speech, a tendency reinforced by orthography. Similarly to L1 English learners of Spanish described by Schmidt (2014), there are differences between the letter-to-sound correspondences between L1 Hungarian, L2 English, and L3 Spanish. While the letter "z" in both L1 and L2 stands for the voiced alveolar fricative, this is not the case in L3 Spanish. Learners are taught from early on that [z] does not exist in Spanish and "z" either stands for /s/ or /θ/. This explains why learners are slightly more successful at blocking coarticulatory voicing induced by sonorants within the word than across the word boundary. Influences of spelling on both L1 and L2 phoneme acquisition have been attested, but the exploration of this topic is beyond the present study, see Shea (2017) for an overview. Overall, we have seen indications of an effect of the word boundary on the two dynamic phonological processes (RVA and PSV), but the results are process-dependent; more research is needed to investigate the learning of phonological processes lexically vs. in dynamic contexts.

4.3. Individual Patterns

Our results show that perception does not closely mirror production at this stage of acquisition for our participants, but production—as proposed by SLM—is dependent on perception, since the fourth possibility described in Section 1 (the sound/process is not perceived but produced) has not been attested in this study (Table 7). That is, the accurate perception of a process is necessary but not sufficient for its accurate production.

In our experiment, three of the four aforementioned perception–production possibilities (Section 1) have been observed more consistently (Tables 6 and 7), as follows:

1. The sound/process is neither perceived nor produced, i.e., it is not learned at all.
2. The sound/process is perceived but not produced (scenario 1 for Wrembel et al., 2022).
3. The sound/process is both perceived and produced (even if not in a completely target-like way), i.e., it is learned.

One participant (S02) did not perceive and did not produce any of the three laryngeal processes examined. Five learners, although they perceived one or two of the laryngeal processes, did not produce any in a target-like way. None of our participants produced PSV, and only two perceived it. The predominant pattern for RVA and a common pattern for aspiration was that learners perceived them (or to be exact, their absence), but failed to transfer this into their speech productions.

The process most successfully acquired was aspiration, which has frequently been correlated with foreign accent in L2 and L3 speech (see Wrembel, 2015 for an overview). This means that it is a feature easy to notice by native speakers, but also by L2/L3 learners as attested by our study, since these highly proficient learners were mostly aware of the non-target-like VOT realizations. However, the production of this new phonetic category was less unanimous. The overall average was slightly below the cut-off value (40 ms); thus, **H4** (*aspiration in English is likely to be attested in the speech of these advanced learners*) is not supported by our data. Only two learners approximated native-like VOT durations (S05 and S11), while another four participants moved away from L1 patterns and produced in-between values. This is also in line with earlier studies (e.g., Flege, 1991), and, again, corroborates the idea that the two modalities are not aligned. It also suggests that in production, the finetuning of a phonetic category is easier than the acquisition of dynamic alternations.

Unfortunately, there are no recent phonetic studies on the VOT values of voiceless stops in Hungarian, but it is interesting to note that the average values in Gósy and Ringen (2009) are considerably lower than in Juhász (2024), despite a methodological difference that would suggest an opposite finding. While Gósy and Ringen (2009) measured the interval from the release burst till the appearance of F2, Juhász (2024) only segmented VOT from the burst to the appearance of the quasi-periodic wave form, and neither of the studies details how the beginning was determined (this methodological issue regarding VOT measurements was also raised in Section 2.2.2). This might mean that there is perhaps an ongoing shift in Hungarian towards longer VOT values, in which case the values in this study may correspond to L1 rather than to hybrid realizations. The longer VOT values might also be attributed to L1 attrition induced by L2. Further acoustic experiments will have to clarify this with monolingual speakers. Juhász (2024) tested L1 Hungarian Mandarin Chinese language learners, and although she controlled for language mode, L2 intrusion into L1 cannot be excluded either.

Individual differences, among them metalinguistic awareness, have long been acknowledged to contribute to successful foreign language learning. An interesting finding of this study is that metaphonological awareness (Section 3.5) is a predictor of accurate perception, but not necessarily of accurate production (although accurate production is dependent on accurate perception). Only two participants (S11 and S12) perceived all three processes as linguistically relevant, and only these two participants had a metaphonological awareness score over 10 points. No individual differences have been identified in the "non-linguistic", purely phonetic perception of PSV (Section 3.3), but their linguistic interpretation correlates with metaphonological awareness. Further research needs to tease apart which aspects of metaphonological knowledge contribute to the identification of postlexical processes and phonetic categories, and what kind of knowledge or skills are necessary to transfer perception into production.

5. Conclusions

The present study has demonstrated that in non-contrastive phonological processes and phonetic category formation, perception and production are not aligned even in the interlanguage of advanced learners, although accurate production depends on accurate perception. Metaphonological awareness seems to influence perception but does not directly influence production. Phonetic category formation (once noticed) is easier to learn (or at least to approximate) than mastering dynamic phonological processes, since in this case, word-level encoding is sufficient, while dynamic processes seem to be more similar to procedural knowledge, and, thus, require more practice. Our findings also suggest that L1 postlexical processes are very difficult to overcome. Some limitations of the study are that it does not include L1 benchmark data from the participants and that it does not examine the potential interaction of the languages involved. Further studies should explore the interaction between markedness and L1/L2/L3 in postlexical phonological processes, and to identify what factors are necessary to transfer perception into production.

Author Contributions: Conceptualization, Z.B. and Z.G.K.; methodology, Z.B. and Z.G.K.; software, Z.B. and Z.G.K.; validation, Z.B. and Z.G.K.; formal analysis, Z.B. and Z.G.K.; investigation, Z.B. and Z.G.K.; resources, Z.B. and Z.G.K.; data curation, Z.B. and Z.G.K.; writing—original draft preparation, Z.B. and Z.G.K.; writing—review and editing, Z.B. and Z.G.K.; visualization, Z.G.K.; supervision, Z.B. and Z.G.K.; project administration, Z.B.; funding acquisition, Z.B. and Z.G.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Hungarian Research, Innovation and Development Office Grant K142498.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and in line with the ethical practice of Eötvös Loránd University, Budapest (Approval Code: PZ1216/22; Approval Date: 1 October 2022).

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

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Acknowledgments: We would like to thank the four anonymous reviewers and the editors for their invaluable contributions. We are also grateful to Anna Hamp and Firas Shbeeb for their help with the acoustic experiments as well as to all the participants of the experiments.

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

Appendix A

Test sentences of the production experiment

English RVA sentences:

Environment	Test word	Sentence
Internal	baseball	His baseball bat was lost by the airline.
Internal	baseball	The white baseball cap is my favourite.
Internal	crossbar	The crossbar was hit by the ball.
Internal	disbelief	Our disbelief was enormous at the news.
Internal	misbehaved	Kate misbehaved during classes.
Internal	disdain	His disdain for pop music was well-known.
Internal	disgust	Her disgust explained everything.
Internal	misgiving	My only misgiving is that I don't trust him.
Final	atlas	The atlas burned in the fire.
Final	bonus	The bonus deal made everybody happy.
Final	campus	The campus guard closed the gates.
Final	famous	That famous game contains hard puzzles.

English VOT sentences (all test words were the sentence-initial words):

Sentence
Cable TV is becoming less popular.
Cancel your plans for tomorrow.
Carry an umbrella in case of rain.
Coffee is Jenny's favourite drink.
Panic is a basic human emotion.
Paying with cash is not allowed.
Tactic may change over time.
Talent can lead to great success.

Spanish PSV sentences:

Environment	Test word	Sentence
Internal	Bosnia	En Bosnia entierran a prisioneros.
Internal	Islandia	En Islandia hay erupciones cada poco.
Internal	Yasmin	La vida de Yasmin está llena de amor.
Internal	asma	Provoca asma infantil en jóvenes.
Internal	asno	Ese asno domado llevaba sombrero.
Internal	cosmos	Nuestro cosmos se expandía lentamente.
Internal	desleal	Fue desleal a su propio país.
Internal	desmontar	Vamos a desmontar sus malvados planes.
Internal	dislexia	La dislexia se diagnostica hoy en día.
Internal	esnobismo	Aquel esnobismo cultural se nota.
Internal	fantasma	Ve el fantasma de la repetición electoral;
Internal	fantasmas	Despertó fantasmas del pasado.
Internal	isla	En esta isla tropical hallará playas.
Internal	islam	El islam recupera su protagonismo.
Internal	mismo	Esa misma noche la robaron.
Internal	plasma	El plasma se movía bastante rápido.
Internal	trasladar	Intenta trasladar su sabiduría.
Final	bolsos	Tiene bolsos listos para el congreso.
Final	campus	Es campus líder en innovación.
Final	casas	Hay casas lindas con jardines.
Final	casas	Las casas modernas se venden bien.
Final	coches	Fabrica coches míticos de marca.
Final	coches	Pasan coches limpios por aquí.

Texts of the perception experiments

English:

Kate had a mischievous little housedog named Tom. This dog loved to misbehave and his carelessness often caused trouble. He was a cute little puppy, people called him a nice boy but he had a habit of jumping on the postman, who really disliked him and called him a disgusting dog. One day Tom thought to himself, "let's go and discover the basement!". As he went down, he heard a sound coming from a corner. He cautiously approached the area and found a house mouse looking back at him with frightened eyes. Tom's first instinct was to pounce on the mouse but something in his little face made him hesitate. As he watched the mouse, he realized that he was just like him—a curious little animal. Tom slowly backed away, not wanting to scare him anymore. He left the basement feeling proud and happy to have made a new friend.

Spanish:

Cuenta una vieja leyenda guaraní que Yasí, la diosa luna, hace muchísimo tiempo quiso conocer la tierra y ver con sus propios ojos lindos todas las maravillas que apenas podía ver entre la espesura de la selva, allá abajo. Un día con su amiga, Araí, la diosa nube, bajaron a la tierra en la forma de dos jóvenes bellísimas acompañadas de dos cisnes bonitos. Por la noche, cansadas de recorrer todo y maravillarse, buscaron un lugar donde descansar. Vieron una isla y en la isla una cabaña. Un atisbo de luz se colaba por la puerta. Cuando se dirigían hacia ella, un jaguar saltó sobre ellas, pero al momento cayó herido. Se arriesgó y las liberó Tupá, el cazador. Como regalo por su valentía, Yasí empezó a sembrar semillas mágicas de un plato con esmalte verde. Desde entonces los guaraníes toman yerba mate.

Metaphonological awareness test

Kérjük, válaszoljon az alábbi kérdésekre. ("Please answer the following questions.")[4]

1. Jelölje be a hanghatárokat (pl. d | a | r | a | b) ("Indicate the sound boundaries.")
    a. madár
    b. vissza
    c. mindnyájan
    d. különb
    e. látja
2. Húzza alá azokat a szótagokat, amelyek ugyanarra a hangra végződnek, mint a *roncs* szó. ("Underline those syllables that end in the same sound as the word *roncs*.")
    a. vezessen
    b. kibonts
    c. kincses
    d. korcstól
    e. bridzshez
3. Jelölje meg a 'z' hangokat az alábbi szavakban. ("Indicate the sound 'z' in the following words.")
    a. szikrázik
    b. kézhez
    c. részben
    d. Oslo
    e. meztelen

Jelölje meg a helyes válaszokat. ("Mark the correct answers.")

4. A spanyol 'p' hang. . .
    a. olyan, mint a magyar 'p' hang
    b. olyan, mint az angol 'p' hang

c. jobban hasonlít a magyar ‘p’ hanghoz, de nem ugyanolyan
d. jobban hasonlít az angol ‘p’ hanghoz, de nem ugyanolyan

(“The Spanish ‘p’ sound is. . .
like the Hungarian ‘p’ sound
like the English ‘p’ sound
more similar to the Hungarian ‘p’ sound, but not the same
more similar to the English ‘p’ sound, but not the same”)

5. A spanyol ‘e’ hang a *café* szóban. . .

a. olyan, mint a magyar ‘e’ hang
b. olyan, mint a magyar ‘é’ hang
c. a magyar ‘e’ és ‘é’ hangok között van
d. leginkább az angol *bad* szó magánhangzójához hasonlít

(“The Spanish ‘e’ sound in the word *café* is. . .
like the Hungarian ‘e’ sound
like the Hungarian ‘é’ sound
between the Hungarian ‘e’ and ‘é’ sound
is more similar to the vowel of the English word *bad*”)

6. A spanyol ‘e’ hang a *coche* szóban. . .

a. olyan, mint a magyar ‘e’ hang
b. olyan, mint a magyar ‘é’ hang
c. a magyar ‘e’ és ‘é’ hangok között van
d. leginkább az angol *bad* szó magánhangzójához hasonlít

(“The Spanish ‘e’ sound in the word *coche* is. . .
like the Hungarian ‘e’ sound
like the Hungarian ‘é’ sound
between the Hungarian ‘e’ and ‘é’ sound
is more similar to the vowel of the English word *bad*”)

7. A spanyol ‘ñ’ hang. . .

a. olyan, mint a magyar ‘ny’ hang, és általában ugyanott is fordul elő a szavakban
b. olyan, mint a magyar ‘ny’ hang, de általában nem ugyanott fordul elő a szavakban
c. olyan, mint az angol ‘n’ hang az *onion* szóban
d. sem az angol ‘n’ (*onion*), sem a magyar ‘ny’ hanggal nem egyezik

(“The Spanish ‘ñ’ sound is
like the Hungarian ‘ny’ sound and usually occurs in the same position within words
like the Hungarian ‘ny’ sound, but is usually not found in the same positions
like the English ‘n’ sound in the word *onion*
not the same as either the English ‘n’ (onion) or the Hungarian ‘ny’ sounds”)

8. Az angol szóhangsúly. . .

a. nagyon hasonló a magyar szóhangsúlyhoz
b. hasonló a magyar szóhangsúlyhoz, csak nagyobb hatással van a magánhangzókra
c. nagyon hasonló a spanyol szóhangsúlyhoz
d. hasonló a spanyol szóhangsúlyhoz, csak nagyobb hatással van a magánhangzókra

(“English word stress is. . .
very similar to Hungarian word stress
similar to Hungarian word stress, but has a greater influence on vowels
very similar to Spanish word stress
similar to Spanish word stress, but with a greater influence on vowels”)

Notes

[1] The experiment focused on the effect of minimal pairhood on voicing production, comparing the voicing of /s/ and /z/ in words that form minimal pairs with other words that do not. Here, we ignore this distinction and report the means of all the words.

[2] A statistical analysis yielded a significant difference between the means of /k/ and /p/ and /k/ and /t/, but the difference between the means of /p/ and /t/ was not significant.

[3] A reviewer indicated that an AX task with shorter stimulus interval would have been more adequate. This is a valid point, though the results suggest that we managed to collect data as intended.

[4] The English translations of the instructions are included here in parentheses for convenience; they were not part of the original test sheets.

References

Abramson, A. S., & Whalen, D. (2017). Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics*, *63*, 75–86. [CrossRef]

Amengual, M. (2021). The acoustic realization of language-specific phonological categories despite dynamic cross-linguistic influence in bilingual and trilingual speech. *The Journal of the Acoustical Society of America*, *149*(2), 1271–1284. [CrossRef]

Bates, D., Maechler, B. B. M., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Bárkányi, Z. (2014). Reflexiones sobre la asimilación de sonoridad de la /s/. In Z. Fábián, I. Szijj, I. Szilágyi, & B. Déri (Eds.), *GPS 60°: Köszöntő kötet Giampaolo Salvi 60. születésnapjára. Studi di linguistica neolatina per i 60 anni di Giampaolo SalviGPS 60* (pp. 25–36). ELTE.

Bárkányi, Z. (2018). The acquisition of voicing assimilation by advanced Hungarian learners of Spanish. *Revista Española de Lingüística Aplicada/Spanish Journal of Applied Linguistics*, *31*(1), 1–31. [CrossRef]

Bárkányi, Z., & G. Kiss, Z. (2021). The perception of voicing contrast in assimilation contexts in minimal pairs: Evidence from Hungarian. *Acta Linguistica Academica*, *68*(1–2), 207–229. [CrossRef]

Bárkányi, Z., & G. Kiss, Z. (2023). Production and perception of voicing contrast in assimilation contexts in Hungarian. *Journal of Uralic Linguistics*, *2*(1), 5–49. [CrossRef]

Bárkányi, Z., & G. Kiss, Z. (2024). Learning and unlearning voicing assimilation. *Frontiers in Language Sciences*, *2*, 1304666. [CrossRef]

Beckman, J., Jessen, M., & Ringen, C. (2013). Empirical evidence for laryngeal features: Aspirating vs. true voice languages. *Journal of Linguistics*, *49*(2), 259–284. [CrossRef]

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In H. C. Nusbaum (Ed.), *The devvelopment of speech perception: Theoretical and methodological issues in cross-language speech research* (pp. 171–204). York Press.

Best, C. T., & Tyler, M. D. (2007). Commonalities and complementarities: Nonnative and second-language speech perception. In O.-S. Bohn, & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil flege* (pp. 13–34). John Benjamins Publishing Company. [CrossRef]

Bialystok, E., & Ryan, E. B. (1985). Towards a definition of metalinguistic skill. *Merrill-Palmer Quarterly*, *31*, 229–251.

Boersma, P., & Weenink, D. (2022). *Praat: Doing phonetics by computer* (Version 6.2.23) [Computer program]. Available online: http://www.praat.org (accessed on 8 October 2022).

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, *22*(3), 129–159. [CrossRef]

Christensen, R. H. B. (2022). *Ordinal: Regression models for ordinal data* (R package version 2022.11-16). Available online: https://github.com/runehaubo/ordinal (accessed on 16 November 2022).

Dmitrieva, O., Jongman, A., & Sereno, J. A. (2020). The effect of instructed second language learning on the acoustic properties of first language speech. *Languages*, *5*(4), 44. [CrossRef]

Draxler, C., & Jänsch, K. (2004). *SpeechRecorder: A universal platform independent multi-channel audio recording software*. Available online: https://www.bas.uni-muenchen.de/Bas/software/speechrecorder/ (accessed on 29 November 2022).

Eckman, F. R. (1985). Some theoretical and pedagogical implications of the markedness differential hypothesis. *Studies in Second Language Acquisition*, *7*(3), 289–307. [CrossRef]

Escalante, C. (2016). Intervocalic voicing and regressive voicing assimilation in L2 Spanish /s/. *Divergencias. Revista de Estudios Lingüísticos y Literarios*, *14*, 30–45.

Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization* [Ph.D. thesis, University of Utrecht].

Flege, J. E. (1982). Laryngeal timing and phonation onset in utterance-initial English stops. *Journal of Phonetics*, *10*(2), 177–192. [CrossRef]

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*(1), 47–65. [CrossRef]

Flege, J. 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*(1), 395–411. [CrossRef]

Flege, J. E. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issue in cross-language research* (pp. 157–232). York Press.

Flege, J. E., & Bohn, O.-S. (2021). The revised speech learning model (SLM-r). In *Second language speech learning theoretical and empirical progress* (pp. 3–83). Cambridge University Press. [CrossRef]

Gósy, M. (2000). A [p, t, k] mássalhangzók zöngekezdési ideje. *Magyar Nyelvőr*, *124*, 195–204.

Gósy, M., & Ringen, C. (2009, August 30–September 1). *Everything you always wanted to know about VOT in Hungarian*. 9th International Conference on the Structure of Hungarian (ICSH9), Debrecen, Hungary.

Gráczi, T. E., & Kohári, A. (2014). Multiple bursts in Hungarian voiceless plosives and VOT measurements. In S. Fuchs, M. Grice, A. Hermes, L. Lancia, & D. Mücke (Eds.), *Proceedings of the 10th international seminar on speech production (ISSP)* (pp. 158–161). University of Cologne.

Hahn, E. D., & Soyer, R. (2005). Probit and logit models: Differences in the multivariate realm. *The Journal of the Royal Statistical Society Series B*, *67*(2), 1–12.

House, A. S., & Fairbanks, G. (1953). The influence of consonantal environment upon the secondary acoustical character-istics of vowels. *Journal of the Acoustical Society of America*, *25*, 105–113. [CrossRef]

Hualde, J. I. (2005). *The sounds of Spanish*. Cambridge University Press.

James, C. (1999). Language awareness: Implications for the language curriculum. *Language, Culture and Curriculum*, *12*(1), 94–115. [CrossRef]

Jansen, W. (2004). *Laryngeal contrast and phonetic voicing: A laboratory phonology approach to English, Hungarian, and Dutch* [Ph.D. dissertation, University of Groningen].

Juhász, K. (2024). *A mandarin beszédhangok produkciója kínaiul tanuló magyar anyanyelvűek ejtésében* [Ph.D. dissertation, Eötvös Loránd Tudományegyetem].

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, *60*(2), 286–319. [CrossRef]

Keating, P. A., Westbury, J. R., & Stevens, K. N. (1980). Mechanisms of stop-consonant release for different places of articulation. *The Journal of the Acoustical Society of America*, *67*(S1), S93–S93. [CrossRef]

Kluender, K. R., Diehl, R. L., & Wright, B. A. (1988). Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics*, *16*(2), 153–169. [CrossRef]

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2020). *lmerTest: Tests in linear mixed effects models* (R package version 3.1-3). Available online: https://github.com/runehaubo/lmerTestR (accessed on 15 November 2022).

Lenth, R. V. (2023). *Emmeans: Estimated marginal means, aka least-squares means* (R package version 1.8.4-1). Available online: https://github.com/rvlenth/emmeans (accessed on 15 November 2022).

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *WORD*, *20*(3), 384–422. [CrossRef]

Liu, J., & Lin, J. (2021). A cross-linguistic study of L3 phonological acquisition of stop contrasts. *Sage Open*, *11*(1), 1–15. [CrossRef]

Lock, J. L. (1972). Ease of articulation. *Journal of Speech and Hearing Research*, *15*(1), 194–200. [CrossRef]

Lüdecke, D., Ben-Shachar, M. S., Makowski, D., Waggoner, P., Patil, I., & Wiernik, B. M. (2021). *Performance: Assessment of regression models performance* (R package version 0.10.4). Available online: https://easystats.github.io/performance/ (accessed on 15 November 2022).

R Core Team. (2022). *R: A language and environment for statistical computing; version 4.2.2. R Foundation for Statistical Computing*. Available online: https://www.R-project.org (accessed on 15 November 2022).

Schmidt, L. B. (2014). Contextual variation in L2 Spanish: Voicing assimilation in advanced learner speech. *Studies in Hispanic and Lusophone Linguistics*, *7*(1), 79–113. [CrossRef]

Schmidt, L. B., & Willis, E. W. (2011). Systematic Investigation of Voicing Assimilation of Spanish /s/ in Mexico City. In S. M. Alvord (Ed.), *Selected proceedings of the 5th conference on laboratory approaches to romance phonology* (pp. 1–20). Cascadilla Proceedings Project.

Shea, C. (2017). L1 English/L2 Spanish: Orthography–phonology activation without contrasts. *Second Language Research*, *33*(2), 207–232. [CrossRef]

Shuster, L. I., & Cottrill, C. (2015). Ease of articulation: A replication. *Journal of Communication Disorders*, *56*, 1–7. [CrossRef] [PubMed]

Siptár, P., & Törkenczy, M. (2000). *The phonology of Hungarian*. Oxford University Press.

Skarnitzl, R., & Šturm, P. (2017). Voicing assimilation in Czech and Slovak speakers of English: Interactions of segmental context, language and strength of foreign accent. *Language and Speech*, *60*(3), 427–453. [CrossRef] [PubMed]

Sypiańska, J. (2013). *Quantity and quality of language use and attrition of L1 Polish due to L2 Danish and L3 English* [Ph.D. thesis, Adam Mickiewicz University].

Szigetvári, P. (2020). Emancipating lenes A reanalysis of English obstruent clusters. *Acta Linguistica Academica*, *67*(1), 39–52. [CrossRef]

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., . . . Yutani, H. (2019). Welcome to the Tidyverse. *The Journal of Open Source Software*, *4*(43), 1686. [CrossRef]

Wrembel, M. (2015). *Search of a new perspective: Cross-linguistic influence in the acquisition of third language phonology* [Seria filologia angielska 48]. Adam Mickiewicz University. Available online: http://hdl.handle.net/10593/14647 (accessed on 4 July 2024).

Wrembel, M., Gut, U., Kopečková, R., & Balas, A. (2020). Cross-linguistic interactions in third language acquisition: Evidence from multi-feature analysis of speech perception. *Languages*, *5*(4), 52. [CrossRef]

Wrembel, M., Gut, U., Kopečková, R., & Balas, A. (2022). The relationship between the perception and production of L2 and L3 rhotics in young multilinguals; an exploratory cross-linguistic study. *International Journal of Multilingualism*, *21*(1), 92–111. [CrossRef]

Young-Scholten, M. (2004). Prosodic constraints on allophonic distribution in adult L2 acquisition. *International Journal of Bilingualism*, *8*(1), 67–77. [CrossRef]