

Article

Cue Weighting in Perception of the Retroflex and Non-Retroflex Laterals in the Zibo Dialect of Chinese

Bing Dong¹, Jie Liang¹ and Chang Liu^{2,*}

¹ Department of English, School of Foreign Languages, Tongji University, Shanghai 200092, China; dongbingtj@tongji.edu.cn (B.D.); liangjie56@163.net (J.L.)

² Department of Speech, Language, and Hearing Sciences, The University of Texas at Austin, Austin, TX 78712, USA

* Correspondence: changliu@austin.utexas.edu

Abstract: This study investigated cue weighting in the perception of the retroflex and non-retroflex lateral contrast in the monosyllabic words /l̥ə/ and /l̥ə/ in the Zibo dialect of Chinese. A binary forced-choice identification task was carried out among 32 natives, using computer-modified natural speech situated in a two-dimensional acoustic space. The results showed that both acoustic cues had a significant main effect on lateral identification, with F1 of the following schwa being the primary cue and the consonant-to-vowel (C/V) duration ratio as a secondary cue. No interaction effect was found between these two acoustic cues. Moreover, the results indicated that acoustic cues were not equally weighted in production and perception of the syllables /l̥ə/ and /l̥ə/ in the Zibo dialect. Future studies are suggested involving other acoustic cues (e.g., the F1 of laterals) or adding noise in the identification task to better understand listeners' listening strategies in their perception of the two laterals in the Zibo dialect.

Keywords: retroflex and non-retroflex laterals; the schwa following the laterals; Zibo dialect; F1; consonant-to-vowel (C/V) duration ratio; cue weighting



Citation: Dong, B.; Liang, J.; Liu, C. Cue Weighting in Perception of the Retroflex and Non-Retroflex Laterals in the Zibo Dialect of Chinese. *Behav. Sci.* **2023**, *13*, 469. <https://doi.org/10.3390/bs13060469>

Academic Editor: Scott D. Lane

Received: 19 April 2023

Revised: 31 May 2023

Accepted: 31 May 2023

Published: 4 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Laterals are common in languages across the world, existing in about 82% of the 317 sample languages in the UPSID (UCLA Phonological Segment Inventory Database). Among these laterals, about 87% are produced in the dental/alveolar region, while the retroflex lateral only appears in 6.7% of them [1]. The comparative rarity of the phonemic retroflex lateral vs. dental/alveolar lateral contrast has motivated research on different languages, for example the Dravidian languages of India, the indigenous languages of Australia and several dialects of Chinese.

Research on the retroflex and non-retroflex lateral contrast in the Dravidian languages spoken in South Asia (e.g., India, Pakistan, Sri Lanka) includes articulatory and acoustic studies of Tamil [2–4], Malayalam [5–7] and Kannada [7]. The articulatory studies on the retroflex and non-retroflex lateral contrast in the Dravidian languages found that there was a dental/alveolar vs. retroflex lateral contrast with a subapical articulation of the retroflex lateral in Tamil [2–4], Malayalam [5–7] and Kannada [7], while previous studies of the Australian indigenous languages Arrernte, Pitjantjatjara and Warlpiri have showed that the retroflex lateral had an apical articulation [8,9].

In addition to articulation measures, acoustic features such as F1, F2 and F3 as well as the duration of laterals have also been examined in previous studies. No consistency was found in the F1 or F2 frequencies between the retroflex and non-retroflex laterals in the Dravidian languages [2–7]. Similarly, in the Australian indigenous languages Arrernte, Pitjantjatjara and Warlpiri, no significant difference was found in the F1 or F2 between the two laterals [9–11]. As for F3, there was a consistently lower F3 in the retroflex lateral /l̥/ than its non-retroflex counterpart /l/ in Tamil, Malayalam and Kannada as well as

in the Australian indigenous languages Arrernte, Pitjantjatjara and Warlpiri. Moreover, the duration of the retroflex lateral /ɭ/ in the Dravidian languages mentioned above was found to be shorter than that of the non-retroflex /l/. For the three Australian indigenous languages, Arrernte, Pitjantjatjara and Warlpiri, no significant difference was found in duration between the two laterals. In sum, a significantly lower F3 in /ɭ/ than in /l/ was found in the Dravidian languages and the three Australian indigenous languages, which was also considered the main correlate of retroflexes in previous studies of vowels and consonants [12–14].

Aside from the articulatory and acoustic characteristics of the retroflex lateral, the phonotactics of /ɭ/ and /l/ have also been discussed in previous research. The occurrence of retroflexes within a syllable and a prosodic word is cross-linguistically asymmetrical. In general, retroflex consonants in CV structure were neutralized or not as well distinguished as in VC structure [8,14–16]. In both the Dravidian languages of India and the indigenous languages of Australia, the retroflex lateral /ɭ/ was found to frequently occur intervocally, less often in the word-final position and rarely in the word-initial position [2,7–11].

Although a number of studies have investigated retroflex and non-retroflex laterals in multiple languages, little research has been carried out on the production and perception of the two laterals in the sound systems of Chinese dialects, such as the Zibo dialect, which are unique and different from other languages. The retroflex and non-retroflex lateral contrast (/ɭ/ vs. /l/) in the Zibo dialect of Chinese differed in both its acoustic and phonotactic characteristics from the lateral contrast in other languages of the world. Therefore, these two laterals in the Zibo dialect were selected as the speech stimuli in this study of lateral perception.

According to the phonology of the Zibo dialect, the retroflex lateral can only appear in the syllable-initial position followed by an atypical schwa, which is higher and only appears following the retroflex lateral /ɭ/ as in /ɭə/, while the non-retroflex lateral /l/ also only appears in the syllable-initial position, but can be followed by most vowels in the Zibo dialect, including monophthongs, diphthongs and triphthongs [17].

In addition, acoustic studies have indicated that the F1 frequency of /ɭ/ in the Zibo dialect is slightly but significantly lower than that of /l/, while no significant difference for F2 and F3 frequencies was found between the two laterals [18,19], consistent with the findings of research on laterals in several other dialects of Chinese [20,21]. Interestingly, the formant structures of the schwa following the two laterals in the Zibo dialect showed similarities and differences. That is, the F1 frequency of the schwa following /ɭ/ was significantly lower than that following /l/ with little speaker variability, while there was no significant difference in F2 frequency of the following vowels [18,19]. The F3 frequency for the schwa following /ɭ/ was significantly higher than that for the schwa following /l/, but the difference was small with large speaker variability [19]. Moreover, the duration of the retroflex lateral /ɭ/ was significantly longer than the non-retroflex lateral /l/, and the duration of the schwa following /ɭ/ was significantly shorter than that following /l/. Therefore, the retroflex lateral /ɭ/ had a significantly larger consonant-to-vowel (C/V) duration ratio in the monosyllabic word /ɭə/ than the non-retroflex lateral /l/ in /lə/. These acoustic measures indicated that the F1 frequency of the schwa following the lateral and the C/V duration ratio were two primary acoustic production cues separating the two laterals in the Zibo dialect and were selected for acoustic manipulation for the perceptual measures in this study.

The C/V duration ratio, defined as the consonant closure duration divided by the preceding vowel duration, is a primary acoustic cue for voicing in several western languages such as English and German [22–26]. For example, it plays a critical role in English voicing in syllable-final and intervocalic positions, e.g., his vs. hiss and dibber vs. dipper, when other cues such as closure voicing and aspiration are ambiguous [22,23]. It is also employed as a voicing cue in speech production and perception for several other Germanic languages, such as Swedish and German [24–26]. On the other hand, the role of the C/V duration ratio in the identification of laterals is still unknown, although acoustic studies

have indicated its importance in separating the retroflex and non-retroflex laterals in production in the Zibo dialect [18,19]. In comparison with absolute durations of the vowel and consonant segments that are highly dependent on speaking rate, the C/V duration ratio, largely independent of speaking rate, is a significantly more powerful and consistent cue for phonological voicing in English [23]. Thus, in the present study, the C/V duration ratio rather than the absolute duration of speech segments was selected as a primary acoustic factor to manipulate. Since the laterals in the Zibo dialect only appear in the initial position of a CV structure, the C/V duration ratio in this study is defined as the duration ratio of the lateral with respect to the following vowels, as in the monosyllabic words /lə/ and /lə/.

Overall, the two laterals in the Zibo dialect are unique due to its phonological structure (e.g., only at the syllable initial position) and the significant differences shown in two acoustic cues: F1 frequency of the schwa following the laterals and the C/V duration ratio of laterals. However, relatively few studies have been performed on the perception of retroflex vs. non-retroflex laterals, which contrast in a sound system such as Zibo Chinese. Thus, this study was to investigate how the two important acoustic cues affected perception of the words /lə/ and /lə/ in the Zibo dialect. The two acoustic cues were manipulated systematically at five levels each, resulting in a total of 25 stimuli. Two research questions were proposed: first, whether and how each acoustic cue influenced the identification of the two laterals; and second, what the perceptual weight of each acoustic cue was, e.g., the role of the other acoustic cue when one acoustic cue was ambiguous.

2. Methods

2.1. Participants

A group of 32 Zibo natives (17 females and 15 males) took part in this study. All participants were aged between 25 and 43 years old, with normal hearing and speaking ability, who communicated in the Zibo dialect in their daily lives. Listeners were paid for their participation.

2.2. Stimuli

As mentioned above, in addition to the F1 frequency of the following vowel and the C/V duration cue, the two laterals also differed in the F1 frequency by themselves. A pilot experiment with four young native speakers of the Zibo dialect suggested that the schwa following the lateral was dominant to determine the identification of laterals. That is, the identification of a hybrid stimulus (e.g., the initial /l/ mixed with the /ə₂/ following /l/ to generate /lə₂/ or the initial /l/ mixed with the /ə₁/ following /l/ to generate /lə₁/) was determined by the following vowel rather than the initial lateral. For example, /lə₂/ was identified as /lə₂/ 100%, while /lə₁/ was perceived as /lə₁/ 100%. Thus, in this study, only two acoustic cues—the F1 frequency of the following vowel and C/V duration ratio—were manipulated, but not the F1 frequency of the initial lateral.

All stimuli were computer-modified natural speech situated in a two-dimensional acoustic space. The original audio materials used in this study were extracted from the natural utterances of one female native speaker (aged 36), recorded in a soundproof recording booth using a Sennheiser GSP 602 headset microphone at a sampling rate of 44.1 kHz. The word list for this study contained 80 compound words with the two laterals in the Zibo dialect, namely 40 words each for /lə/ and /lə/. The monosyllabic words /lə/ and /lə/ in the word list had the same tone: a low falling tone with a tone value of 31 in the Zibo dialect, which is the most frequent citation tone where /lə/ and /lə/ appear in this dialect. They were recorded by the female speaker with a Chinese carrier sentence /və tu [compound] tə lə sā piā/ or /və tu [compound] tə lə sā piā/ (English meaning: “I read /lə/ as in [compound] three times” or “I read /lə/ as in [compound] three times”). The carrier sentence was designed to help the speaker pronounce more naturally [27], with a sentence focus on the monosyllabic word /lə/ or /lə/ (the second /lə/ or /lə/ in the carrier sentence). The monosyllabic words /lə/ and /lə/, rather than those in the compounds, were extracted for stimulus generation. Altogether 40 monosyllabic /lə/ words and 40

monosyllabic /lə/ words were recorded and extracted from the carrier sentence. Based on the judgement of two phoneticians, one /lə/ and one /lə/ were selected from the recordings that were clearly pronounced and had no creakiness in the syllable pronunciation. The two syllables were equalized in duration, room-mean-square level and pitch contour, and then used as the standard signal to synthesize speech stimuli with the manipulation of the F1 frequency of the following schwa and the C/V duration ratio in Praat [28].

Since the lateral and the following schwa have a one-to-one match in the two words, two sets of speech stimuli with systematic changes in the two acoustic cues were generated in this study: one set based on /lə₁/ and the other set based on /lə₂/ . Here, [ə₁] and [ə₂] are used respectively to distinguish between the schwas following /l/ and /l/ for ease of readability.

In the first set of stimuli, the standard signal /lə₁/ (with equal duration, intensity, and pitch contour with the standard signal /lə₂/ used for the synthesis of the second set of stimuli) was used to generate the speech stimuli. As shown in Table 1, first, the F1 value of the vowel segment [ə₁] in /lə₁/ was manipulated at five levels based on the average F1 values of 40 tokens of [ə₁] and 40 tokens of [ə₂] from the recording, e.g., 661 Hz for [ə₁] and 840 Hz for [ə₂]. The five F1 frequencies were equally distant on the auditory Bark scale. Second, the C/V duration ratio continuum was also created at five levels with the average C/V duration ratio of the 40 recordings of /lə₁/ as level 1 and the average C/V duration ratio of the 40 recordings of /lə₂/ as level 5. For the speech synthesis, except for the changes in the two acoustic cues manipulated above, other acoustic features such as the F2 and F3 frequencies were not changed. Altogether, there were a total of 25 different stimuli (5 levels on the F1 frequency × 5 levels on the C/V duration ratio) for the first set of speech stimuli, which used /lə₁/ as the standard signal.

Table 1. Levels of F1 value of [ə₁] (Cue 1) and levels of the consonant-to-vowel (C/V) duration ratio of the lateral (Cue 2) of the stimuli.

(a)					
Levels of Cue 1	F1_1	F1_2	F1_3	F1_4	F1_5
F1 of /ə/ (Hz)	661	702	746	791	840
(b)					
Levels of Cue 2	DR_1	DR_2	DR_3	DR_4	DR_5
C/V duration ratio	1.36	1.64	1.92	2.20	2.48
Duration of C (ms)	131	139	147	154	162
Duration of V (ms)	96	88	80	73	65

Likewise, another set of speech stimuli with similar manipulations of the two acoustic cues, e.g., five levels on the F1 of the following schwa and five levels on the C/V duration ratio, were generated with the syllable /lə₂/ as the standard signal. Altogether, 50 different stimuli (25 stimuli generated from /lə₁/ and 25 stimuli generated from /lə₂/) were synthesized in the study. In sum, for the two sets of stimuli, only the F1 frequency of the following schwa and the C/V duration ratio of the lateral were changed systematically, with all other acoustic features unchanged.

The perception experiment was carried out in a quiet room on a computer with Sennheiser headphones (Model 280 Pro) using ExperimentMFC implemented in Praat. Subjects were required to perform a binary forced-choice word identification task in which they clicked on one of two Chinese characters presented on the computer screen, i.e., /lə/ with the corresponding Chinese character meaning “two” or /lə/ with the corresponding Chinese character meaning “hot”, in a self-paced and self-selected fashion after hearing a speech signal. Once listeners responded, the next trial was automatically started. The listeners had up to three chances to listen to a single stimulus with no response time limit. For each listener, the 50 speech stimuli were presented in a random order in one block and a similar block was repeated three times, resulting in a total of 150 stimulus presentations.

The formal experiment of each participant was preceded by an eight-trial practice session to familiarize the subjects with the procedure. No feedback was provided in the practice or test sessions.

2.3. Data Analysis

Data collected from Praat were separated into two groups: data for 75 stimuli generated from /l̩₁/ (5 formant levels × 5 duration ratio levels × 3 repetitions) and data for 75 stimuli generated from /l̩₂/ (5 formant levels × 5 duration ratio levels × 3 repetitions). The data were analyzed separately in order to see the results of perception with one acoustic cue in the “lateral-schwa” structure fixed and the other cue changing along the continuum. A mixed-effects logistic regression model was used for the stimuli of each group to quantify the use of the two cues simultaneously by modeling how well category affiliation was predicted by the F1 value of the following schwa or the C/V duration ratio in the syllable. The logistic regression model tests for significant main effects of the two cues and interaction between cues were used to evaluate differences in cue weight across conditions [29]. The general logistic regression equation for the present identification task is given as the following formula:

$$\log(\text{odds} (/l̩_1/)) = \ln(p(/l̩_1/)/p(/l̩_2/)) = \alpha + \beta_{F1} \times \text{F1 value of schwa} + \beta_{DR} \times \text{the C/V duration ratio of lateral} + \beta_{\text{Interaction}} \times (\text{F1 value of schwa} \times \text{the C/V duration ratio of lateral})$$

In this equation, α is the intercept of the regression model. The coefficients (β s) of the logistic regression model quantify the impact of a one-step difference in one of the cues on the log odds of a subject's response. Morrison [30] has suggested that these coefficients can be interpreted as a measure of a subject's reliance on each cue. In other words, the magnitude of the coefficient reflects the degree to which subjects use a specific cue in making their responses.

3. Results

3.1. Results for Stimuli Generated from /l̩₁/

In the response percentages results for /l̩₁/ on different levels of Cue 1 (F1 value of the following schwa), as expected, the subjects' judgments vary systematically with the increase in the F1 value of the following schwa (from 661 Hz in step 1 to 840 Hz in step 5), as seen in the left panel of Figure 1. Generally, the /l̩₁/ response curve to stimuli generated from /l̩₁/ tends to decrease with Cue 1 changing from level 1 to level 5, which suggests that Cue 1 has a large effect on participants' responses to the stimuli. That is, the lower the F1 value of the following schwa, the more likely they are identified by the participants as /l̩₁/.

In terms of the results for percent /l̩₁/ responses for different levels of Cue 2, i.e., the C/V duration ratio in the syllable “lateral-schwa”, the average percentage of /l̩₁/ response increased gradually from step 1 to step 5. As seen in the right panel of Figure 1, the average percentage of /l̩₁/ responses increased gradually at the beginning from step 1 at 63.13% to 72.29% at step 5, which indicates a general tendency that for all stimuli on different levels of Cue 2, the percentage of /l̩₁/ responses tends to increase with Cue 2 changing from level 1 to level 5. Generally, the results suggest that for the stimuli in this identification task, the larger the C/V duration ratio, the more likely the participants identified the stimuli as /l̩₁/.

To evaluate the effects of the two cues on the identification of the monosyllabic words /l̩/ and /l̩₂/, a mixed-effects logistic regression model was used to analyze the data. Cue 1 and Cue 2 were set as fixed effects and Participant and Stimuli were set as random effects. The analysis of the model shows that the marginal R^2 is 0.754 and the conditional R^2 is 0.872, indicating that 75.4% of the variance is caused by the independent effects in the model (Cue 1 and Cue 2), 11.8% of the variance is due to the random effects (Participant and Stimuli) and the model can explain a total of 87.2% of the variance. The results of the mixed-effects logistic regression model are shown in Table 2.

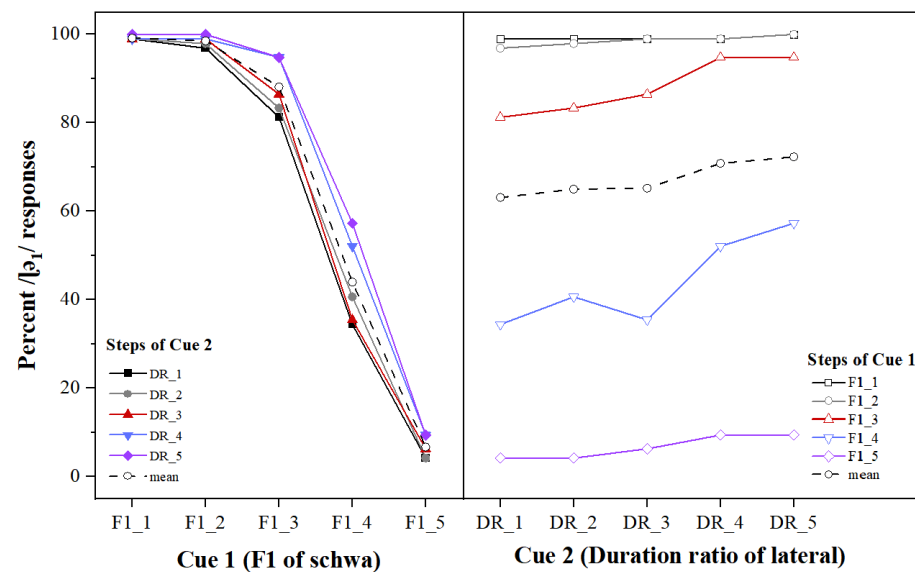


Figure 1. Percentages of / ə_1 / response as a function of Cue 1 (left panel) and Cue 2 (right panel) for stimuli generated from / ə_1 /, with colored solid lines representing five different levels and the black dashed line representing the mean.

Table 2. Coefficients of the two cues in the mixed-effects logistic regression model for stimuli generated from / ə_1 /.

Predictor: Fixed Effects	β	SE	Exp (B)	z	p
Intercept	2.78	0.34	16.16	8.27	<0.001
Cue 1 (F1)	−3.17	0.15	0.04	−20.83	<0.001
Cue 2 (C/V)	0.50	0.09	1.65	5.56	<0.001
Cue 1 $\times$ Cue 2	−0.13	0.08	0.88	−1.75	0.080

Results show that first, there is a significant main effect of Cue 1 ($\chi^2(1) = 434.07$, $p < 0.001$), indicating that an increase in the F1 frequency steps of the following schwa decreases / ə / responses; and second, there is a significant main effect of Cue 2 ($\chi^2(1) = 30.96$, $p < 0.001$), indicating that an increase in the steps of the C/V duration ratio increases / ə / responses. No significant interaction effect was found between Cue 1 and Cue 2 ($p = 0.080$).

These results suggest that subjects used both cues in their identification of the stimuli. The coefficients of the logistic regression analysis show to what extent a one-step difference in one of the cues caused a change in the log odds of a participant's response of / ə_1 / . As for cue weighting, Cue 1 ($\beta = -3.17$) is the primary cue for identification of the syllable / ə / and Cue 2 ($\beta = 0.50$) is the secondary perceptual cue, suggesting that Zibo natives weight the F1 value of the following schwa much more heavily than the C/V duration ratio in the syllable on their identification of / ə / and / ə_1 /.

3.2. Results for Stimuli Generated from / ə_2 /

For the results of response percentages for / ə_2 / on different levels of Cue 1, the subjects' judgments varied with the increase in the F1 value of the following schwa, as shown in the left panel of Figure 2.

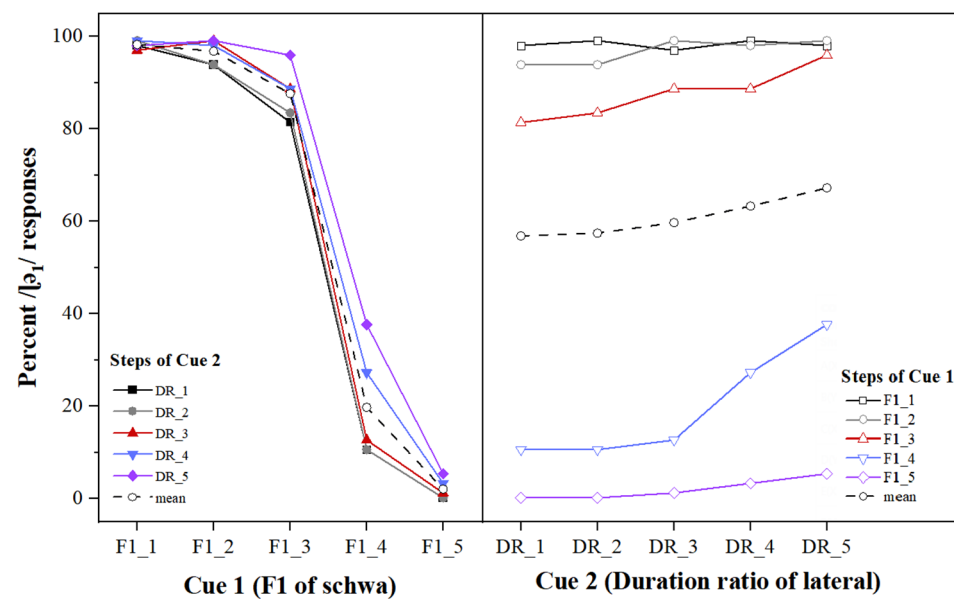


Figure 2. Percentages of / ə_1 / response as a function of Cue 1 (left panel) and Cue 2 (right panel) for stimuli generated from / ə_2 /, with colored solid lines representing five different levels and the black dashed line representing the mean.

These systematic decreases show that participants relied on Cue 1 heavily in their responses to the stimuli. For stimuli on different Cue 1 levels, generally the lower the F1 frequency of the following schwa is, the higher the percentage of / ə_1 / response is, which is identical to the result for stimuli in the first set. For the results for percentage of / ə_1 / responses on different levels of Cue 2, the average percentage of / ə_1 / response increased systematically from 56.67% to 67.08% from step 1 to step 5 of the C/V duration ratio, as seen in the right panel of Figure 2.

A mixed-effects logistic regression model was used to evaluate the effect of the two cues on the identification of / ə / and / ə / for this group of stimuli. Similarly, Cue 1 and Cue 2 were set as fixed effects, and Participant and Stimuli were set as random effects. The analysis of the model shows that the marginal R^2 is 0.720 and the conditional R^2 is 0.880, showing that 72% of the variance is caused by the independent effects in the model (Cue 1 and Cue 2), 16% of the variance is due to the random effects (Participant and Stimuli) and the model can explain a total of 88% of the variance. The results of the mixed-effects logistic regression model are shown in Table 3.

Table 3. Coefficients of the two cues in the mixed-effects logistic regression model for stimuli generated from / ə_2 /.

Predictor: Fixed Effects	β	SE	Exp (B)	z	p
Intercept	1.18	0.41	3.27	2.88	0.004
Cue 1	−3.07	0.24	0.05	−12.96	<0.001
Cue 2	0.51	0.20	1.66	2.59	0.009
Cue 1×Cue 2	0.17	0.15	1.19	1.11	0.266

The results show that first, there is a significant main effect of Cue 1 ($\chi^2(1) = 168.06$, $p < 0.001$), indicating that an increase in the steps of Cue 1 decreases / ə / responses; and second, there is a significant main effect of Cue 2 ($\chi^2(1) = 6.73$, $p = 0.009$), indicating that an increase in the steps of Cue 2 increases / ə / responses. No significant interaction effect was found between Cue 1 and Cue 2.

These results indicate that subjects use both of the two cues in their identification of the stimuli of this set. Cue 1 ($\beta = -3.07$) is the primary cue for the identification of the syllable / ə / and Cue 2 ($\beta = 0.51$) is the secondary perceptual cue, suggesting that Zibo natives

weight the F1 value of the following schwa much more heavily than the C/V duration ratio in the identification of the two laterals, which echoes the findings in Section 3.1.

4. Discussion

The purpose of this study was to assess how the two acoustic cues, i.e., the F1 frequency of the following schwa and the consonant-to-vowel duration ratio, affected the perception of the two laterals in the monosyllabic words /l_ə1/ and /l_ə2/. The two acoustic cues were systematically manipulated to generate two sets of speech stimuli, one originated from /l/ and the other originated from /l/.

4.1. Cue Weighting of the Two Acoustic Cues in the Perception of /l_ə1/ and /l_ə2/

For the two sets of stimuli, both the F1 of the following schwa and the C/V duration ratio in the monosyllabic words had a significant main effect on the identification of the two laterals. That is, the lower the F1 of the following schwa, the greater the possibility it was identified as /l_ə1/, while the higher the F1 of the following schwa, the higher possibility it was identified as /l_ə2/. As for the C/V duration ratio of the lateral in the “lateral-schwa” syllable, in general, the larger the duration ratio of the lateral in the syllable was, the more likely it was identified as /l_ə1/; the smaller the C/V duration ratio of the lateral in the syllable was, the more likely it was identified as /l_ə2/. No interaction effect was found between the two acoustic cues in either set of stimuli. The results from the mixed effects logistic regression model showed that the β of the F1 of the following schwa was much larger than that of the C/V duration ratio, suggesting that the F1 frequency of the following schwa was the primary cue, while the C/V duration ratio played a secondary role in the identification of the two laterals. In particular, as shown in Figures 1 and 2, when the F1 frequency of the following schwa dominated phonetic identification, the change in the C/V duration ratio had little effect on the identification of the two laterals, also indicating the critical role of the F1 frequency of the following schwa.

In fact, spectral information as the primary cue with phonetic duration as the secondary cue has also been found in a number of previous studies. For example, F1 and F2 frequencies are the primary cues for English vowel perception, while vowel duration plays a secondary role for native speakers [31,32]. In the present study, since duration is not a distinctive phonemic feature in the Zibo dialect, the effect of speech duration may be even smaller, if any.

As mentioned previously, the C/V duration ratio was used in the present study to reflect the temporal information of the lateral in relation to that of the following schwa in the monosyllabic word. When the spectral information was ambiguous (e.g., see the curves of Cue 1 at level 3 and 4 in Figures 1 and 2), this temporal information significantly affected the identification as a secondary cue. This result is similar to the finding in Port and Dalby [23] on the role of the C/V duration ratio in the perception of voiced and voiceless consonants in the intervocalic and final positions, in which although the C/V duration ratio might be dominated by non-temporal cues, it showed a significant impact on the perception of syllable-final voicing in English when other acoustic features were held constant and yet ambiguous.

4.2. Production and Perception Link of Laterals in the Zibo Dialect

Contrasts for speech sounds differ along multiple phonetic dimensions and phonetic cue weighting can be quantified in the context of both production and perception. Research examining the relative alignment of cue weighting across modalities has revealed both parallels and asymmetries between the modalities [29]. For example, English speakers’ productions of syllable-initial /b/ and /p/ showed large and consistent differences in VOT. In addition, /b/ was followed by a lower f0 than /p/ on average, but such differences were much smaller and less consistent than those in VOT [33,34]. This difference in cue weighting between the two cues in production is also reflected in perception: when asked to categorize sounds varying in the duration of aspiration and f0, subjects’ responses mainly depended on aspiration (the primary cue), with f0 in a secondary but detectable role [35]. This is an example of a match of both cue use and cue weighting between production

and perception of a phonemic contrast. Languages differ in the use of phonetic cues and the relative importance given to these cues. Cue use in perception tends to reflect community production norms on a broad level; however, research work directly comparing individual use across modalities has revealed matches and mismatches in cue use between perception and production. Shultz et al. [36] explored the connections between production and perception by investigating the manner in which native English speakers' relative weighting of VOT corresponded to that of onset f0 in both production and perception of English syllable-initial consonant voicing. Their finding showed a significant negative correlation of VOT and onset f0 in production, and the coefficients suggested that all the participants were primarily users of VOT, with some participants putting remarkable weight on onset f0 and other participants not using onset f0 in their production. However, in perception, no statistically significant results were found for the correlation between the two cues, although with a positive trend. These results indicated that the relative weighting of acoustic cues might be processed differently in production and perception of the same phonemic contrast.

This study also showed both a match and a mismatch in acoustic cue use between the production and perception of /l̥₁/ and /l̥₂/ in the Zibo dialect. The two important cues that showed large and consistent differences between /l̥₁/ and /l̥₂/ in production [18,19], i.e., the F1 of the following schwa and the C/V duration ratio of the laterals, were also important in perception. However, native speakers' production of /l̥₁/ and /l̥₂/ showed large and consistent differences in both the F1 value of the following schwa and the C/V duration ratio of the lateral, while in perception, participants' responses were mainly determined by the F1 value of the following schwa (the primary cue), with the C/V duration ratio as a secondary cue. This suggests that the two cues differed in their importance in production and perception, showing a mismatch of cue weighting between these two modalities. In speech production, speakers aim to produce an integrated acoustic property to instantiate the contrast between two speech units while compensating for idiosyncratic variation [36]. In contrast, in perception, people may use one cue primarily with other cues as secondary. The asymmetry in cue weighting between the production and perception of Zibo laterals may be associated with the different strategies people use during speech production and perception.

4.3. Limitation and Future Studies

It is worth mentioning that the percentages of /l̥₁/ responses in both groups of stimuli were over chance rate (50% in this study), showing a tendency of more /l̥₁/ responses than /l̥₂/ in both sets of stimuli. There are two possible reasons. First, this result is possibly related to the word frequency effect in word processing, in which high-frequency words are processed more efficiently than low-frequency words [37,38]; secondly, when listeners are unsure of what they have heard, they are more likely to report hearing a high-frequency lexical item than a low-frequency one [39]. Therefore, the word exposure difference might lead to the favorable response bias toward /l̥₁/ in identification due to its higher frequency than /l̥₂/ when used both independently and in compound words in the daily life of Zibo natives. Second, the higher /l̥₁/ responses for stimuli based on /l̥₁/ than for stimuli based on /l̥₂/ also indicates that aside from the F1 frequency of the following schwa as well as the C/V duration ratio of the lateral in the syllable, other acoustic cues, such as the F1 frequency of the laterals and the higher formants of laterals and schwas, may also affect the identification of the syllables. In addition, the manipulation of lateral and vowel durations in this study may result in a change in the naturalness of formant transition, and the effect of such a change needs further investigation in future study.

During speech perception, listeners must decide which cues are relevant and determine the relative importance of each cue as well as integrate other signal-external cues [29]. The C/V duration ratio of the lateral may seem much less important than the formant cue in the identification of the two Zibo laterals, but it may facilitate successful perception of speech under adverse listening conditions, e.g., in the presence of interfering noise and

when listeners have non-native linguistic backgrounds or hearing impairments. In future studies, stimuli involving other acoustic cues (e.g., a lower F1 in the retroflex lateral) in different listening conditions (e.g., quiet and noisy) can be used to investigate listeners' cue-weighting strategies in their perception of the two laterals in the Zibo dialect in a variety of listening environments.

5. Conclusions

In the present study, two acoustic cues, the F1 of the schwa following the lateral and the C/V duration ratio, were systematically manipulated to investigate their roles in the identification of two laterals: /l/ and /ɭ/ in the Zibo dialect in Chinese. Both acoustic cues showed significant effects on lateral perception. Moreover, listeners relied on the F1 of the following schwa more heavily than the C/V duration ratio in the syllable. These results suggest that although both cues were presented consistently by native speakers in their production, listeners had different listening strategies in cue weighting for their perception.

Author Contributions: Conceptualization, B.D. and J.L.; methodology, B.D. and C.L.; software, B.D.; validation, B.D. and J.L.; formal analysis, B.D. and C.L.; writing—original draft preparation, B.D.; writing—review and editing, B.D. and C.L.; visualization, B.D.; supervision, C.L. and J.L.; project administration, J.L. and C.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of the School of Foreign Languages of Tongji University (protocol code: tjsflrec202212, approved in 2022).

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

1. Maddieson, I. *Patterns of Sounds*; Cambridge University Press: Cambridge, UK, 1984.
2. McDonough, J.; Johnson, K. Tamil liquids: An investigation into the basis of the contrast among five liquids in a dialect of Tamil. *J. Int. Phon. Assoc.* **1997**, *27*, 1–26. [\[CrossRef\]](#)
3. Narayanan, S.S.; Kaun, A. Acoustic modeling of Tamil retroflex liquids. In Proceedings of the 14th International Congress of the Phonetic Sciences, San Francisco, CA, USA, 1–7 August 1999; pp. 2097–2100.
4. Narayanan, S.S.; Byrd, D.; Kaun, A. Geometry, kinematics, and acoustics of Tamil liquid consonants. *J. Acoust. Soc. Am.* **1999**, *106*, 1993–2007. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Punnoose, R.; Khattab, G.; Al-Tamimi, G. The contested fifth liquid in Malayalam: A window into the lateral-rhotic relationship in Dravidian languages. *Phonetica* **2013**, *70*, 274–297. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Scobbie, J.M.; Punnoose, R.; Khattab, G. Articulating five liquids: A single speaker ultrasound study of Malayalam. In *Rhotics: New Data and Perspectives*; Spreafico, L., Viette, A., Eds.; Bozen-Bolzano University Press: Bolzano, Italy, 2013; pp. 99–124.
7. Tabain, M.; Kochetov, A. Acoustic realization and inventory size: Kannada and Malayalam alveolar/retroflex laterals and /ɭ/. *Phonetica* **2018**, *75*, 85–109. [\[CrossRef\]](#)
8. Tabain, M.; Kochetov, A.; Beare, R. An ultrasound and formant study of manner contrasts at four coronal places of articulation. *J. Acoust. Soc. Am.* **2020**, *148*, 3195–3217. [\[CrossRef\]](#)
9. Tabain, M.; Beare, R. An ultrasound study of coronal places of articulation in Central Arrernte: Apicals, laminals and rhotics. *J. Phonetics* **2018**, *66*, 63–81. [\[CrossRef\]](#)
10. Tabain, M.; Butcher, A.; Breen, G.; Beare, R. An acoustic study of multiple lateral consonants in three Central Australian languages. *J. Acoust. Soc. Am.* **2016**, *139*, 361–372. [\[CrossRef\]](#)
11. Tabain, M.; Butcher, A.; Breen, G.; Beare, R. A formant study of the alveolar versus retroflex contrast in three Central Australian languages: Stop, nasal, and lateral manners of articulation. *J. Acoust. Soc. Am.* **2020**, *147*, 2745–2765. [\[CrossRef\]](#)
12. Ladefoged, P. *Elements of Acoustic Phonetics*, 2nd ed.; The University of Chicago Press: Chicago, IL, USA, 1996.
13. Stevens, K.N. *Acoustic Phonetics*; The MIT Press: Cambridge/London, UK, 1998.
14. Hamann, S. *The Phonetics and Phonology of Retroflexes*; LOT: Utrecht, The Netherlands, 2003.
15. Steriade, D. Positional Neutralization and the Expression of Contrast. Master's Thesis, University of California, Los Angeles, CA, USA, 1994.

16. Ohala, M.; Ohala, J.J. Acoustic VC transitions correlate with degree of perceptual confusion of place contrast in Hindi. *Trav. Cercle Linguist. Cph.* **2001**, *31*, 265–284.
17. Meng, Q.; Luo, F. *Records of Zichuan Dialect*; Yuwen Press: Beijing, China, 1994.
18. Dong, B.; Liang, J. Retroflex and non-retroflex laterals in the Zibo dialect: An experimental case study. In Proceedings of the 22nd Conference of the Oriental COCOSA, Cebu City, Philippines, 25–27 October 2019.
19. Dong, B. Production and Perception of the Retroflex and Non-Retroflex Laterals in Zibo Dialect of Chinese. Ph.D. Thesis, Tongji University, Shanghai, China, 2023.
20. Zhu, X.; Jiao, N. The retroflex lateral approximant in Jincheng dialect: An intermediate stage in the development of Er. *Nankai Linguist.* **2006**, *7*, 33–39.
21. Qi, W. A Study on Speech Sounds of Zhongyuan Mandarin in Shandong Province. Ph.D. Thesis, Shandong University, Jinan, China, 2019.
22. Denes, P. Effect of duration on the perception of voicing. *J. Acoust. Soc. Am.* **1955**, *27*, 761–764. [\[CrossRef\]](#)
23. Port, R.F.; Dalby, J. Consonant/vowel ratio as a cue for voicing in English. *Percept. Psychophys.* **1982**, *32*, 141–152. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Elert, C.C. *Phonological Studies of Quantity in Swedish*; Uppsala Almqvist & Wiksell: Stockholm, Sweden, 1964.
25. Kohler, K.J.; Kunzel, H.J. *The Temporal Organization of Closing-Opening Movements for Sequences of Vowels and Plosives in German*; Arbeitsberichte; University of Kiel: Kiel, Germany, 1978.
26. Kohler, K.J. Dimensions in the perception of fortis and lenis plosives. *Phonetica* **1979**, *36*, 332–343. [\[CrossRef\]](#)
27. Müller, D. Cue weighting in the perception of phonemic and allophonic laterals along the darkness continuum: Evidence from Greek and Albanian. *Albanohellenica* **2015**, *6*, 1–14.
28. Boersma, P.; Weenink, D. *Praat: Doing Phonetics by Computer*, version 6.2.01; [Computer Program]. Available online: <http://www.praat.org/> (accessed on 18 April 2023).
29. Schertz, J.; Clare, E.J. Phonetic cue weighting in perception and production. *Wires. Cogn. Sci.* **2020**, *11*, E1521. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Morrison, G.S. Logistic regression modelling for first and second language perception data. *Amst. Stud. Theory His.* **2007**, *4*, 219–236.
31. Strange, W.; Jenkins, J.J.; Johnson, T. Dynamic specification of coarticulated vowels. *J. Acoust. Soc. Am.* **1983**, *74*, 695–705. [\[CrossRef\]](#)
32. Hillenbrand, J.M.; Clark, M.J.; Houde, R.A. Some effects of duration on vowel recognition. *J. Acoust. Soc. Am.* **2000**, *108*, 3013–3022. [\[CrossRef\]](#)
33. Lisker, L. “Voicing” in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Lang. Speech* **1986**, *29*, 3–11. [\[CrossRef\]](#)
34. Schertz, J.; Cho, T.; Lotto, A.; Warner, N. Individual differences in perceptual adaptability of foreign sound categories. *Atten. Percept. Psycho.* **2016**, *78*, 355–367. [\[CrossRef\]](#)
35. Abramson, A.S.; Lisker, L. Relative power of cues: F0 shift versus voice timing. *Phon. Linguist. Essays Honor. Peter Ladefoged* **1985**, *15*, 25–33.
36. Shultz, A.A.; Francis, A.L.; Llanos, F. Differential cue weighting in perception and production of consonant voicing. *J. Acoust. Soc. Am.* **2012**, *132*, EL95–E101. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Howes, D. On the relation between the intelligibility and frequency of occurrence of English words. *J. Acoust. Soc. Am.* **1957**, *29*, 296–305. [\[CrossRef\]](#)
38. Monsell, S.; Doyle, M.C.; Haggard, P.N. Effects of frequency on visual word recognition tasks: Where are they? *J. Exp. Psychol. Gen.* **1989**, *118*, 43. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Goldiamond, I.; Hawkins, W.F. Vexierversuch: The log relationship between word-frequency and recognition obtained in the absence of stimulus words. *J. Exp. Psychol.* **1958**, *56*, 457. [\[CrossRef\]](#) [\[PubMed\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.