2025. Proc Ling Soc Amer 10(1). 59128. https://doi.org/10.3765/plsa.v10i1.5912.

A more comprehensive view on the production-perception link of cue weighting in sound change: The case of vowel length contrast in Long'an Zhuang

Yue Yin*

Abstract. This paper aims to provide a more comprehensive view on the production-perception link of cue weighting in sound change by further examining the relationship between the weights of different cues in production and perception, as well as comparing two ways of cue weight calculation (raw vs. relativized). Production and perception experiments were conducted on Long'an Zhuang (a Tai language in Guangxi, China) vowel length contrast, which involves both duration and spectral cues. The results showed a negative correlation between duration and formant cue weights in production but a positive correlation in perception. Additionally, the production-perception link was found to be more robust regarding relative cue weights rather than raw weights. These findings suggest that sound change may be related to asymmetry in production and perception, weakening the production-perception link in absolute cue contribution but not in relative dominance.

Keywords. production-perception link; cue weighting; vowel length; Tai-Kadai

1. Introduction. Phonological contrast usually differs across multiple phonetic cues, which vary in their importance to the phonological contrast. This importance could be conceptualized as cue weight in both production and perception (for an overview, see Schertz & Clare 2020).

Individual differences in cue weighting have been reported in both production (e.g., Shultz et al. 2012; Clayards 2018b; Hauser 2023) and perception (e.g., Hazan & Rosen 1991; Clayards 2018a). However, less is known about the nature of this individual variation in speech perception and production. A common approach to investigating this issue is to examine the relationship between the two domains (Yu 2019). In this context, whether the production-perception link exists has emerged as a critical research question. Although many studies have examined how individual-level production and perception cue weighting are connected, their findings often present inconsistencies (e.g., Shultz et al. 2012; Yu 2019; Schertz et al. 2015), and the underlying reasons for these discrepancies remain unclear.

This paper aims to contribute to research on the production-perception link in cue weighting by investigating an ongoing merger in the vowel length contrast of Long'an Zhuang, a Tai language spoken in Guangxi, China, which relies on both duration and spectral cues. Additionally, to provide a more comprehensive perspective on this topic, the study further examines the relationship between cue weights in production and perception and compares different methods of cue weight calculation.

1.1. Production-perception link. Many theoretical frameworks suggest a potential link between cue weighting in perception and production. For example, Motor Theory (Liberman & Mattingly 1985) and Direct Realism (Fowler 1986) consider gestures as the fundamental units of perception, and thus assume a direct connection between perception and production. A similar prediction could be found in exemplar-based theories (Pierrehumbert 2001), which posits a

* I would like to thank Zihe Li and Baoya Chen from Peking University, Micha Elsner, Becca Morley, Cynthia Clopper, Tianyi Ni, the audience of Phonies at OSU, four reviewers, and the audience at the LSA 2025 Annual Meeting for their constructive feedback. I would also like to thank Hei Vangz from UC Berkeley for her help during the fieldwork. All mistakes are mine. Author: Yue Yin, Ohio State University (yin.1007@buckeyemail.osu.edu).

production-perception loop, where the production and perception share the representation, and the update of representation is based on production output and its feedback to perception.

However, empirical studies have reported inconsistent results, and it remains underexplored why the discrepancy occurs. Harrington et al. (2012) suggested that context of sound variation and change could be a factor, in that the connection between perception and production is weaker when variants are undergoing change or are socially stratified than when they are more stable and not subject to social factors. However, such a hypothesis is not fully supported by empirical data. Although the connection in stable variation has been confirmed in many studies, in vowel nasalization (Beddor et al. 2018; Zellou 2017), fricative contrast (Yu 2019), etc., some other studies (e.g., Shultz et al. 2012) failed to detect it on VOT and f0 cue weighting in English stop voicing.

These uncertainties in previous studies highlight the need for a deeper investigation into the different nature of production and perception individually. Regarding cue weighting, it would be beneficial to compare the relationship between the weights of different cues in production and perception. Additionally, variations in how cue weights are calculated across studies (Schertz & Clare 2020) may also contribute to these inconsistencies.

1.2. RELATIONSHIP BETWEEN THE WEIGHTS OF MULTIPLE CUES. The relationship between the weights of different cues could be described at both token and talker levels (Clayards 2018b; Hauser 2023). Cue relationships at the token level discuss the relationship of cue values across tokens, e.g., the cue trading in speech perception (Repp 1982) and the articulation-based intrinsic link between two cues in production (Clayards 2018b). In contrast, talker-level cue relationships compare the cue weights of multiple cues across different speakers (Hazan & Rosen 1991; Shultz et al. 2012; Schertz et al. 2015). Here, this study focuses on the talker-level relationship.

Analyzing the talker-level cue relationship can provide insights into the sources of individual differences in cue use, which contains at least two possibilities: (1) A positive correlation between cue weights suggests that individual differences lie in overall cue effectiveness (Clayards 2018a). In this case, speakers who are better at using one cue to discriminate the contrast tend to be similarly effective at using the other cue(s). (2) Conversely, a negative correlation indicates a trade-off relationship (Hauser 2023). Specifically, speakers who are more effective at using one cue to distinguish the contrast tend to be less effective at using the other cue(s).

Previous studies have reported inconsistent results in both production and perception. In production, Shultz et al. (2012) found a negative correlation between VOT and f0 cue weights in English stop voicing contrasts, indicating that speakers who produce greater VOT differences tend to produce smaller f0 differences, and vice versa. Similarly, Bang et al. (2018) observed that Korean speakers with higher f0 weights have lower VOT weights, reflecting a cue shift toward quasi-tonogenesis. In contrast, Clayards (2018b) found a positive but non-significant correlation between VOT and f0 in English using Cohens d as cue weights, while LDA coefficients showed a marginally significant correlation.

In perception, some studies found a positive correlation between cue weights, while others did not. For instance, Hazan & Rosen (1991) reported a positive correlation between formant and burst cues (*date-gate*). If the change of one cue makes a greater difference in some speakers' contrast identification, those speakers also tend to be more sensitive to the other cue. Strong positive correlations have also been observed between vowel contrasts (*bet-bat, Luce-lose*) and fricative contrasts (*sock-shock*) (Clayards 2018a). However, for English stop voicing, Shultz et al. (2012) found only a marginally significant positive correlation between VOT and f0, and Clayards (2018a) failed to find a reliable relationship.

In addition, most previous studies have focused on stable phonological contrasts rather than ongoing sound changes, leaving the influence of sound change on cue correlations underexplored. In the sound change literature, while few studies analyze cue weighting correlations at the individual level, many investigations into cue shifts (e.g., tonogenesis) suggest a trade-off relationship in cue weights at the group level (Kuang & Cui 2018; Coetzee et al. 2018; Bang et al. 2018; Lee & Jongman 2018). Additionally, Chen (2022) observed that an ongoing merger showed reduced distinctions for both cues, which may indicate a positive relationship in cue weights. This suggests that sound change may influence cue correlations or, alternatively, that cue correlation patterns could signal the direction of sound change. These findings underscore the importance of examining individual-level cue weight correlations in sound change cases to better understand the micro-level mechanisms of cue weighting.

Overall, the relationship between multiple cue weights remains unclear, with previous studies primarily focusing on consonantal contrasts, especially VOT and F0 in stops, across a limited number of languages. It is also uncertain how this relationship interacts with sound change. This study provides a new test case from the Zhuang vowel length contrast, an ongoing merger (Yin 2024).

1.3. CALCULATING INDIVIDUAL CUE WEIGHTS. To quantify the individual differences of cue weighting in production and perception, studies need a measure of cue weights for each participant (see Schertz & Clare 2020 for an overview).

In production, primary cues are described as having higher "reliability" (Schertz et al. 2015), "precision" (Clayards 2018b), being produced with "larger difference" (Shultz et al. 2012) or more "distinctively" (Hauser 2023). However, there are inconsistencies in how cue weights are calculated in production across previous studies. These approaches can be divided into two categories based on which variables are treated as predictors and which as response variables. The first, classification-based approach, treats acoustic parameters as predictors and focuses on how well they classify categories. For instance, coefficients from Linear Discriminant Analysis (LDA) have been used to quantify cue weights (Shultz et al. 2012; Schertz et al. 2015; Hauser 2023), evaluating each cue's contribution to distinguishing categories. The second approach reverses these roles, treating categories as predictors and acoustic parameters as outcomes. This method quantifies how much each cue varies across categories. For example, Clayards (2018b) used Cohen's d to measure the degree of overlap between categories for each cue, reflecting how precisely the cue conveys contrasts (Clayards 2008). Studies using beta coefficients of the linear regression model (Lee & Jongman 2018) also align with this approach. These two approaches thus reflect different perspectives: how cues predict categories versus how categories explain cue variation. In this study, to make the production and perception more comparable regarding cue weighting, I adopted the first approach to calculate the production cue weights.

In perception, the conceptualization and calculation of cue weights are more consistent across studies. Perceptual cue weight refers to how much each cue contributes to listeners responses in disambiguating different categories, usually through identification or discrimination tasks (Holt & Lotto 2006; Francis et al. 2000). Beta coefficients from logistic regression models are usually used to quantify the cue weights (Morrison 2005; Shultz et al. 2012; Schertz et al. 2015).

Apart from different modeling methods, another key difference in cue weighting studies is whether to relativize cue weights. Many studies use raw values directly (Morrison 2005; Shultz et al. 2012; Schertz et al. 2015), while others relativize coefficients by expressing them as a proportion of the total weight of all cues (Clayards 2018a). These two approaches reflect different aspects of cue use. Raw coefficient values reflect the absolute contribution of a cue to the distinctness or discrimination of the contrast, showing how effectively a speaker uses a specific cue. In contrast, relativized values represent the relative importance of a cue, indicating how much it dominates in production or perception. To address these complementary perspectives, the current study adopts both methods to provide a more comprehensive examination of individual differences in cue weighting.

1.4. THE CURRENT STUDY. Given the uncertainty in previous research on cue weighting in production and perception, this study investigates an ongoing merger in the Long'an Zhuang vowel length contrast as a new test case. Additionally, it offers a more comprehensive perspective on the production-perception link by examining cue correlations in production and perception, and by comparing both raw and relative cue weights.

The research questions are as follows: (1) What is the relationship between the weights of different cues in production and perception? (2) Does the production-perception link manifest in the absolute cue contribution to the contrast or the relative dominance, or both? (3) Does cue weighting in production and perception interact with sound change?

2. Vowel length contrast and the case of Long'an Zhuang.

2.1. DURATION AND SPECTRAL CUES IN VOWEL LENGTH CONTRAST. Multiple acoustic cues of vowel length contrast have been observed across languages (see Lehnert-LeHouillier 2010 for an overview). While duration is usually the primary cue, secondary cues such as spectral differences, f0, and segmental contexts also play a role. Concerning spectral differences, the centralization of short vowels is a typical production pattern cross-linguistically (Maddieson, 1984). Perception studies also showed the primary role of duration as well as the influence of vowel quality in different languages, such as Thai (Abramson & Ren 1990), Dai (Chen 2022), and low vowels in German (Weiss 1974).

Variations of cue use in vowel length contrast have been reported across speakers, especially in sound changes (e.g., Kang et al. 2015; Chen 2022). For example, Chen (2022) found that for the /aːi/-/ai/ pair, younger speakers in Dai have an enhancement of formant distinction than older speakers in production, while both groups use formant as the perceptual primary cue; while for the /aːu/-/au/ pair, both weakening of duration and formant distinction was found, in younger speakers, while duration is the primary cue in perception in both groups. These findings indicate a cue shift for the /aːi/-/ai/ contrast while a merger for /aːu/-/au/. However, the majority of research on vowel length contrast focuses on group-level results, it remains unclear on the patterns of individual-level differences in cue weighting.

2.2. LONG'AN ZHUANG VOWEL LENGTH CONTRAST. Zhuang is known as an important member of the Tai-Kadai languages - Tai branch. The case in the present study is Long'an Zhuang, a Zhuang variety spoken in Long'an County, which is located in the northwest of the capital city Nanning of Guangxi Zhuang Autonomous Region, China, as shown in Figure 1.

Long'an Zhuang has a systematic vowel length contrast involving both duration and formant cues (Zhang et al. 1999; Yin 2024). There are six distinguished vowel qualities /i e a ɯ u ɔ/, with vowel length contrast for all vowels except for /e/. The current study chooses the /aː a/ ([aː ɐ]) pair in stop-final syllables as the target objects. According to Yin (2024), this pair is undergoing

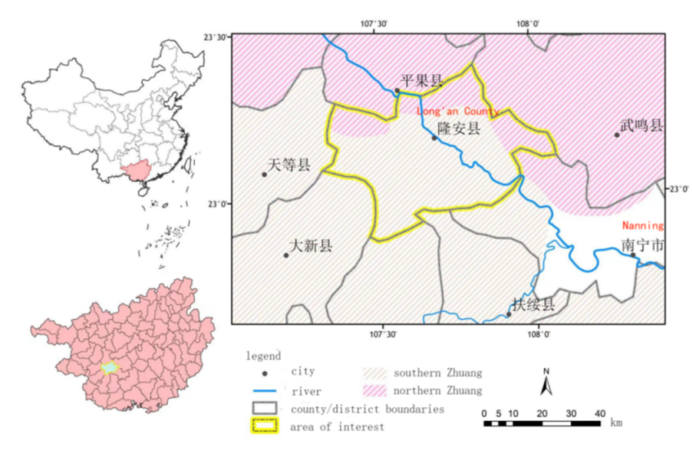

Figure 1. Location of Long'an County

a merging process, which is at the initiation stage. Specifically, this change has been initiated in production, which shows an age effect, but not yet in perception. The experimental data used in this study is the same as that in Yin (2024), but new analyses were done on the data.

3. Production experiment.

3.1. DESIGN. Production data was collected from 34 native speakers (19-71 years old) of Long'an Zhuang, using six target pairs together with fillers. The targets are six stop-final minimal pairs (or near-minimal pairs) of /aː a/ preceding different codas (-p, -t, -k), as shown in Table 1. The fillers include 34 nontarget minimal pairs and 23 consonant+/l/-initial words.

Long	Short
kaːt 55 'break (rope)'	kat 55 'button (clothes)'
paːt 55 'basin'	pat 55 'pen'
laːp 33 'wax'	lap 55 'dark'
taːp 44 'answer'	tap 55 'liver'
paːk 55 'mouth'	pak 55 'north'
t^haːk 213 'sunbathe'	t^hak 21 'break (stick)'

Table 1. Target words of production experiment

The experiment had three sessions, each consisting of the same word list, randomized for each session. Within each session, the word list was presented to the participants visually in their Chinese translations, and the participants were asked to produce each word twice in isolated forms. Before the experiment began, under the guidance of the experimenter, all of the participants first scanned the word list of the first session to ensure that they were familiar with the target vocabulary and could produce the right word. The audios were recorded at a sampling rate of 44,100 Hz through an Audio-Technica AT2020 cardioid condenser microphone connected to a Behringer XENYX 302USB mixing console. The recording software was Audition CC 2023, installed on a Lenovo laptop.

3.2. ACOUSTIC MEASUREMENTS. All target audios were manually segmented and annotated in Praat (Boersma & Weenink 2023), and the acoustic data of vowel duration, F1, and F2 were automatically extracted using Praat scripts and then manually adjusted. 13 tokens were excluded due to misreading or audio quality issues, resulting in 1211 valid data points in total.

The duration of the vowel was measured as the interval between the onset and offset of a clearly visible F2 (Kleber 2020). The formant values were measured as the means of the stable formant section, i.e., the portion that is not undergoing a formant transition. Mean values of the two repetitions of each token were used in the statistical analysis.

4. Perception experiment.

4.1. STIMULI. The target stimuli were synthesized from the naturally produced minimal pair /paːt 55/ 'basin' and /pat 55/ 'pen' of a middle-aged male speaker who didn't participate in the experiments. The duration and formants were modified in incremental steps (2 master sounds $\times$ 5 duration $\times$ 5 formant) using Praat. The original duration and formant values of the sound tokens used for synthesis are shown in Table 2.

Token	Duration (ms)	F1 (Hz)	F2 (Hz)
/paːt 55/ 'basin'	251	863	1431
/pat 55/ 'pen'	171	778	1345

Table 2. The original values of the tokens chosen for synthesis

First, two duration continuua were generated from the original audio of /paːt 55/ and /pat 55/ individually, using the PSOLA algorithm in Praat. The maximum and minimum values were the durations of the original audio of /paːt 55/ and /pat 55/ respectively, and then three points equally spaced between the endpoints were taken to obtain five steps in total. This manipulation resulted in two series of duration continua of two vowel qualities. Then, each long-vowel quality token and its equal-duration counterpart with short-vowel quality were manipulated to form two five-step formant continua with all formants adjusted as a whole: using one token as the master sound and the other as the reference sound, and then the reversing. Each time, the source and the formant values of the master sound, as well as the formant values of the reference sound, were obtained via LPC. The source was then refiltered using formant values at five evenly spaced steps, including the master sound's and reference sound's formant values as the start and end points. The resulting tokens were then normalized for intensity and f0.

Finally, 50 stimuli were obtained (5 duration $\times$ 5 formant $\times$ 2 master sounds), together with 60 fillers, forming 110 stimuli in total. The fillers contained 50 target stimuli for another project, and 5 pairs of naturally produced tokens of everyday vocabulary to ensure the participants' concentration during the experiment.

4.2. PROCEDURE. A two-alternative forced choice (2AFC) identification task was implemented on the same group of participants as those in the production experiment, through Eprime 3.0.3.9. Each stimulus occurred twice, resulting in 220 trials in total. The trials were divided into five blocks, and the participants could take a short rest after each block. In each trial, the audio was presented in isolation and repeated twice. Then, two word meaning options were presented in Chinese characters on the screen, and the participants chose the one they heard. The participants were encouraged to respond both quickly and accurately and not to consider too much. To ensure participants' familiarity with the procedure, 16 practice trials of nasal-final minimal pairs of vowel length contrast were conducted before the experiment. The entire task was completed in 30 to 50 minutes.

5. Data analysis.

5.1. PRODUCTION DATA. Before implementing statistical analysis, I first log-transformed the duration data and then within-speaker z-scored the duration, F1, and F2 data. Then, three separate linear mixed-effects models were done on all participants' data to examine the effect of vowel length (long vs. short) on duration, F1, and F2, using the lmer function lme4 R package (Bates et al. 2015). The random slopes and intercepts on subject and item (minimal pair) were included. Since this package doesn't report p-values for linear mixed-effects models, p-values were calculated by the lmerTest R package (Kuznetsova et al. 2017).

As mentioned in section 1.3, the classification-based approach was adopted to calculate production cue weights here. The coefficients of each linear discriminant (i.e., each cue) from the LDA were calculated as cue weights, which is widely used by previous studies (Shultz et al. 2012; Schertz et al. 2015; Hauser 2023). However, LDA requires the assumption of homoscedasticity (Izenman 2008), which was not met in this data, according to the results of Box's M test (Box's $M = 39.4, p < 0.001$). To ensure that the results of the analysis are not method-dependent, I also employed Random Forest (Breiman 2001, and see Sonderegger & Sskuthy 2025 for application in phonetics) as a comparative approach. Unlike LDA, Random Forest does not rely on assumptions about the data distribution, making it a suitable complementary method for this analysis. The Mean Decrease in Accuracy (MDA) from the Random Forest method was used to quantify cue weights. Instead of estimating a precise relationship between the predictor and the response variable, as is done in traditional data modeling, the measures of variable importance in Random Forest are robust statistics that reflect a variables contribution to the model's simulation of the underlying data-generating process (Archer & Kimes 2008).

Both LDA and Randon Forest models were implemented on each participant's data individually, using the LDA function from the MASS R package (Ripley et al. 2013) and the randomForest function from the randomForest R package (Liaw & Wiener 2002). Based on the raw cue weights, relativized cue weights were then calculated by ratio, i.e., as a proportion of the total of cue weights, to provide an estimate of how much one cue dominates production. Further, to examine the relationship between different cues, pairwise Pearson correlations on duration, F1, and F2 were done on non-relativized cue weights.

5.2. PERCEPTION DATA. For perception data, a mixed-effects logistic regression model was built with the glmer function in the lme4 R package (Bates et al. 2015) on the data of all speakers. The response variable was the perception response (long vs. short), and the fixed effects were duration, formant, and their interaction term. For random effects, the model included random slopes and intercepts on both subjects and master sound used for generating the stimuli. The interaction term between the two predictors was removed in the random effects structure, since the model failed to converge when including it. The model was specified as follows: $response \sim duration * formant + (1 + duration + formant | subject) + (1 + duration + formant | master\_sound)$.

The cue weight of a certain cue of each individual is calculated by adding the fixed effect coefficient and the random slope on that cue for that specific subject. Similar to the treatment of the production data, relativized weights were calculated for further analysis. Pearson correlation analyses were conducted on the non-relativized cue weights of duration and formant to examine the relationships between the two cues.

5.3. PRODUCTION-PERCEPTION LINK. The correlation between production and perception was tested on both non-relativized and relativized cue weights. For non-relativized cue weights, Pearson correlation analyses were conducted between duration in production and perception, F1 in production and formant in perception, and F2 in production and formant in perception. Similarly, Pearson correlation analyses were conducted on relativized cue weights.

6. Results.

6.1. PRODUCTION. As shown in Figure 2, long vowels generally have a longer duration and higher F1 and F2 values, indicating that the short vowels are generally higher and less fronted.

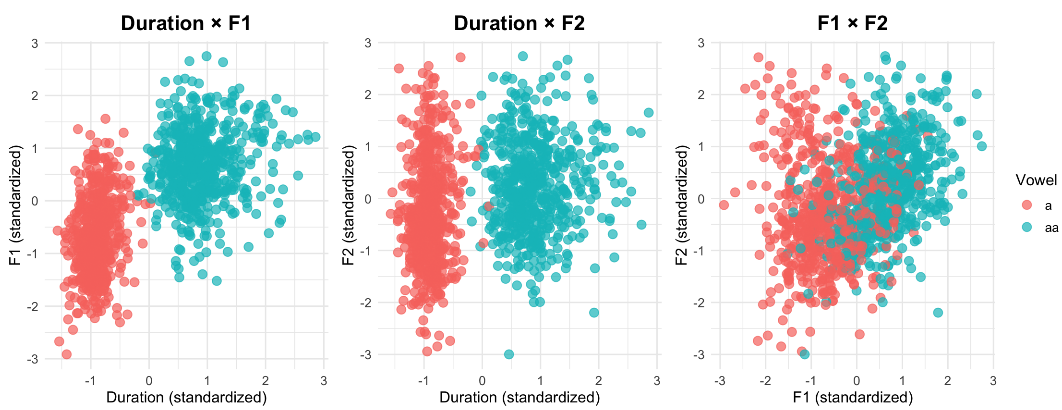

Figure 2. Distribution of acoustic cues (standardized)

This observation is also confirmed by the results of linear mixed-effects models, as shown in Table 3. Significant fixed effects are found on all three dimensions. By comparing the coefficients, it can be concluded that duration is the primary cue at the group level, followed by F1, and then F2.

	β	SE	*df*	t	p
Duration	1.66	0.10	15.37	16.29	< 0.001***
F1	0.58	0.07	12.71	8.26	< 0.001***
F2	0.18	0.07	6.95	2.47	0.043*

Table 3. Results of three linear mixed-effects model on production (fixed effects)

Regarding the correlations between individual cue weights of different cues, a significant negative correlation between duration and F1 was found by both LDA ($R = -0.55, p =< 0.001$) and Random Forest ($R = -0.87, p < 0.001$), as shown in Figure 3 and Figure 4.

6.2. PERCEPTION. As shown by the results of the mixed-effects logistic regression in Table 4, the fixed effects of duration and formant are both significant. This indicates that duration and formant cues both make significant contributions to the identification of the vowel length contrast.

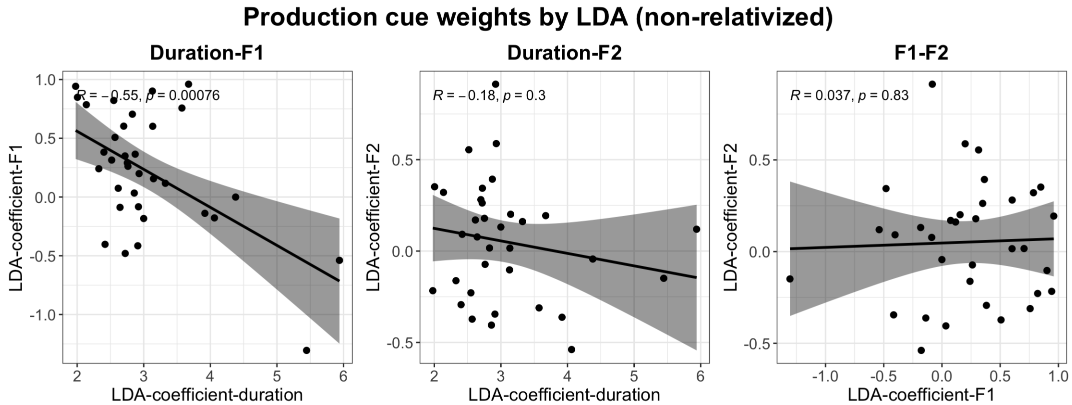

Figure 3. Production cue weights by LDA (non-relativized)

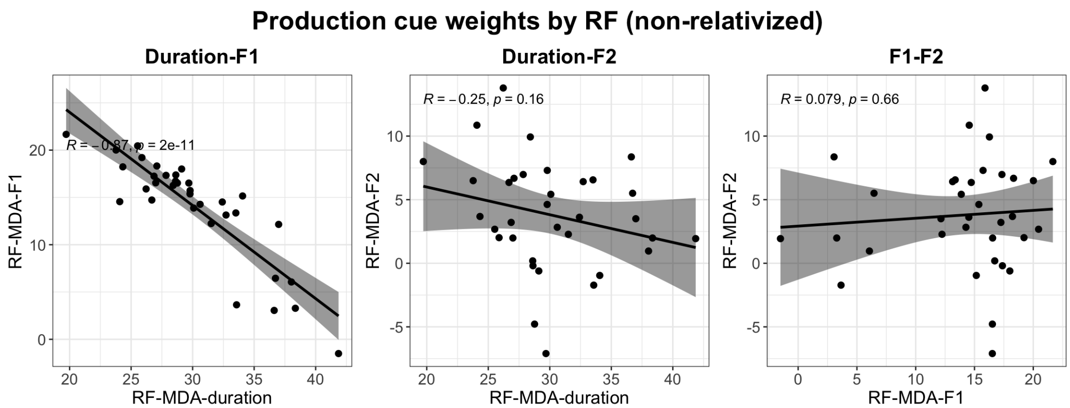

Figure 4. Production cue weights by RF (non-relativized)

Additionally, the beta coefficient of duration is greater than that of formant, suggesting that duration is the primary cue. In addition, the significant positive interaction between duration and formant indicates that the two cues enhance each other's effects.

	β	SE	z	p
(Intercept)	0.72	0.14	5.04	< 0.001***
duration	1.21	0.16	7.37	< 0.001***
formant	0.92	0.20	4.59	< 0.001***
duration: formant	0.19	0.06	3.20	0.001**

Table 4. Results of mixed-effects logistic regression model on perception (fixed effects)

Correlation between individual cue weights in perception shows a different pattern from that in production, in that duration and formant showed a significant positive correlation ($R = 0.60, p < 0.001$), as shown in Figure 5.

6.3. PRODUCTION-PERCEPTION LINK. The production-perception link was examined on both non-relativized and relativized cue weights via correlation analyses. For production data, cue weights calculated by LDA and Random Forest were both included.

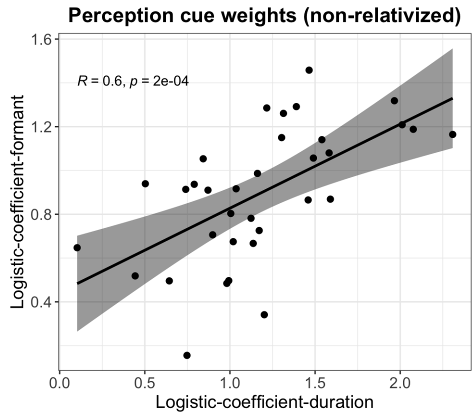

Figure 5. Perception cue weights (non-relativized)

Analyses of non-relativized cue weights reveal slightly differing results between LDA and Random Forest methods. With LDA production weights, significant positive correlations were observed between production and perception cue weights for duration ($R = 0.45, p = 0.008$) and F2 ($R = 0.47, p = 0.005$), as shown in Figure 6. In contrast, with Random Forest production weights, only a significant correlation between production F2 and perception formant cue weights ($R = 0.40, p = 0.019$), as shown in Figure 7. However, since F2 is the least important cue in production, this correlation might not be very informative.

Regarding relativized cue weights, a significant positive correlation between production and perception was detected on duration relativized cue weights, no matter using LDA ($R = 0.40, p = 0.020$) or Random Forest ($R = 0.36, p = 0.036$) on production data, as shown in Figure 8.

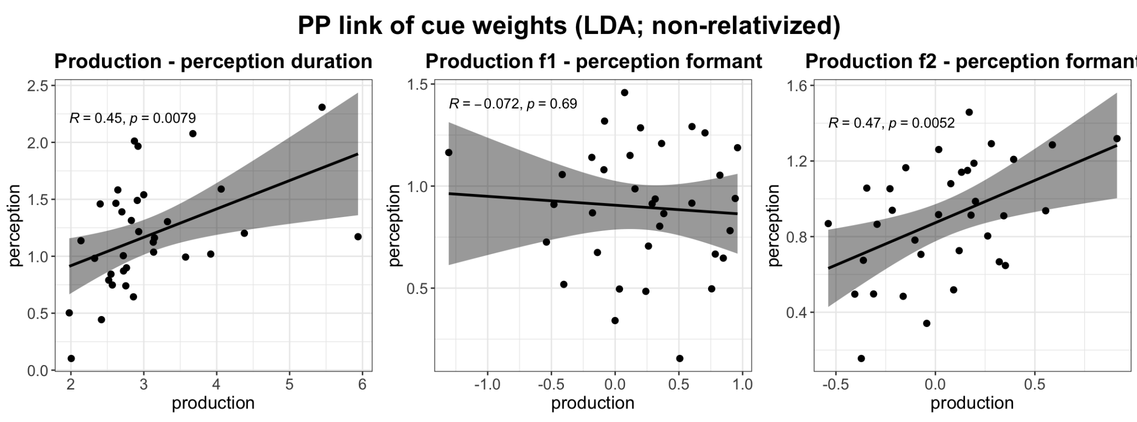

Figure 6. Production-perception link of cue weights (LDA; non-relativized)

7. Discussion. In summary, a significant negative correlation between duration and formant was observed in production, while in perception, a significant positive correlation between cues was found. Regarding the production-perception link, when using LDA to calculate production cue weights, a significant link was found by both raw weights and relativized weights; when using Random Forest, the link was only detected by relativized weights.

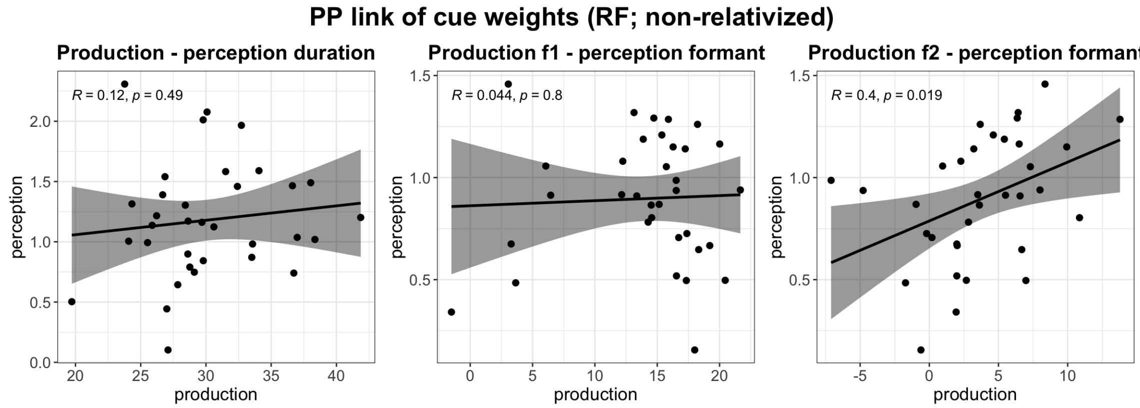

Figure 7. Production-perception link of cue weights (RF; non-relativized)

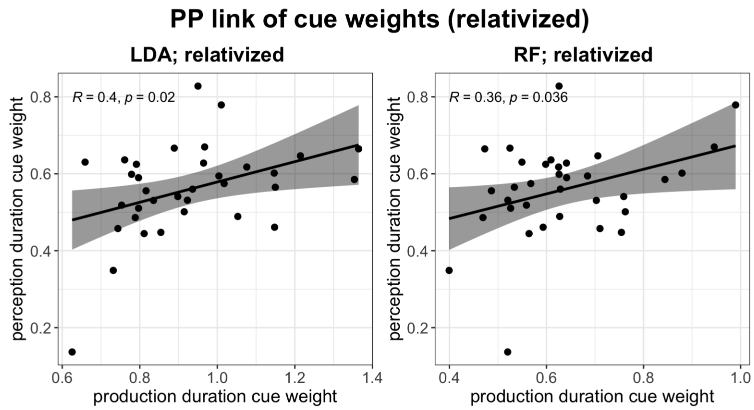

Figure 8. Production-perception link of cue weights (relativized)

7.1. Different patterns in production and perception. In production, the significant negative correlation between cue weights indicates a trade-off relationship, consistent with Shultz et al.'s (2012) findings on English consonant voicing. However, this talker-level relationship differs from token-level "cue trading" (Repp 1982; Howell 1993), which involves perception or production equivalency in sounds with varying cue combinations. In the current study, the trade-off reflects individual differences in cue use: The more a speaker relies on one cue, the less they rely on the other. However, this does not imply equivalency, as the overall ability to distinguish the contrast may vary across individuals.

However, perception shows a different pattern of cue relationship. In perception, a significant positive correlation between cues was found. This suggests that listeners integrate these cues in a parallel manner. This aligns with previous findings in Hazan & Rosen (1991) and Clayards (2018a), indicating that some individuals are better at using phonetic information generally than other individuals. However, the situation could vary from case to case. For example, no significant correlation between VOT and f0 was found in Shultz et al. (2012), and in Clayards (2018a), the author only found significant positive correlations for minimal pairs *bet-bat*, *Luce-lose*, and *sock-shock*, but not for *bog-dog* and *dear-tear*. In the current study, I only examined the /aː a/ pair in stop-final syllables, and more vowel pairs need to be explored to provide a more comprehensive examination.

Overall, these findings show different cue-weighting strategies in production and perception.

7.2. RELATIVE CUE WEIGHTING MATTERS IN PRODUCTION-PERCEPTION LINK. The results of this study reveal a link between the relative cue weights in production and perception. Overall, this finding has confirmed a generally assumed connection between the two domains (e.g., Beddor et al. 2018; Yu 2019). However, this link is more robust on the relativized cue weights, while results on raw cue weights demonstrate a discrepancy across cue weight measures. This result may indicate that the connection between production and perception could show the discrepancy between cue contribution and cue dominance. As noted by Clayards (2018a), raw weights in perception reflect listeners' sensitivity to a specific cue, while relativized weights indicate the extent to which one cue dominates perception. A similar distinction could apply to production.

In addition, the inconsistency between LDA and Random Forest results may reflect differences in the assumptions and applicability of the two methods. Different results between LDA and Cohen's d as production cue weight measures have also been reported (Clayards 2018b) but not under discussion. However, many previous studies relied solely on methods like LDA (Shultz et al. 2012; Schertz et al. 2015) without testing assumptions such as multivariate normality or homoscedasticity, raising concerns about their validity. In contrast, although Random Forest avoids distributional assumptions, it is more typically used for larger datasets with more predictors (Sonderegger & Sskuthy 2025) and may also cause biased results on small datasets, which are common in phonetic experiments. Therefore, faced with the discrepancy across different cue weight measures, comparing multiple methods can yield more robust conclusions and reduce method-dependent biases. Systematic comparisons of these methods could help address current limitations and will be our future direction.

7.3. INTERACTION BETWEEN SOUND CHANGE AND CUE USE. The current findings show a misalignment between the cue correlations of production and that of perception, as well as a more robust production-perception link on cue dominance, indicated by relativized cue weights rather than raw weights. This pattern may be related to the ongoing merger of vowel length contrast in Long'an Zhuang. This sound change has been initiated in production but not yet in perception (Yin 2024), in that some younger speakers may have less distinction in production but with their perceptual sensitivity unaffected, which may be connected to the asymmetry of cue correlations in the two domains.

The failure to detect a robust link on raw weights is consistent with the hypothesis that ongoing sound change could weaken the link between production and perception (Harrington et al. 2012). However, the significant link on relativized weights may indicate that although ongoing sound change could weaken the link of production and perception in absolute cue contribution, the link of relative cue dominance remains.

8. Conclusion. The current study examined the existence of the production-perception link on both cue contribution (raw weights) and cue dominance (relativized weights) in sound change, using the merger of the Long'an Zhuang vowel length contrast as a test case. The talker-level relationships between cues were also tested in production and perception. There are two main findings: First, production and perception could have different patterns regarding correlations between different cues. In this case, production shows a negative correlation, while perception shows a positive correlation between different cues. Second, when facing sound changes, it is possible that the production-perception link is more robust in relative cue dominance than in absolute cue contribution. Overall, these findings provide a more comprehensive view of the production-perception link. Given this, it would be beneficial for future research to further examine the production-perception relationship in more various contexts of sound variation and change, and by comparing different cue weight calculation methods.

References

Abramson, A. S. & N. Ren. 1990. Distinctive vowel length: Duration vs. spectrum in thai. *Journal of Phonetics* 18(2). 79–92.

Archer, Kellie J. & Ryan V. Kimes. 2008. Empirical characterization of random forest variable importance measures. *Computational Statistics & Data Analysis* 52(4). 2249–2260. https://doi.org/10.1016/j.csda.2007.08.015.

Bang, Hye-Young, Morgan Sonderegger, Yoonjung Kang, Meghan Clayards & Tae-Jin Yoon. 2018. The emergence, progress, and impact of sound change in progress in seoul korean: Implications for mechanisms of tonogenesis. *Journal of Phonetics* 66. 120–144. https://doi.org/10.1016/j.wocn.2017.09.005.

Bates, Douglas, Martin Mchler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. https://doi.org/10.18637/jss.v067.i01.

Beddor, P. S., A. W. Coetzee, W. Styler, K. B. McGowan & J. E. Boland. 2018. The time course of individuals perception of coarticulatory information is linked to their production: Implications for sound change. *Language* 94(4). 931–968. https://doi.org/10.1353/lan.2018.0051.

Boersma, Paul & David Weenink. 2023. Praat: doing phonetics by computer (version 6.3.20). `http://www.praat.org/`. [Computer program.] Retrieved October 24, 2023.

Breiman, Leo. 2001. Random forests. *Machine Learning* 45(1). 5–32. https://doi.org/10.1023/A:1010933404324.

Chen, Xiaobei. 2022. *Salient perceptual cues in phonological variation: The decline of vowel length contrast in daiya*. Beijing: Peking University MA thesis.

Clayards, M. 2018a. Differences in cue weights for speech perception are correlated for individuals within and across contrasts. *The Journal of the Acoustical Society of America* 144(3). EL172–EL177. https://doi.org/10.1121/1.5052025.

Clayards, M. 2018b. Individual talker and token covariation in the production of multiple cues to stop voicing. *Phonetica* 75(1). 1–23. https://doi.org/10.1159/000448809.

Clayards, Meghan. 2008. *The ideal listener: Making optimal use of acoustic cues for speech perception*: University of Rochester dissertation.

Coetzee, Andries W., Patrice S. Beddor, Kerby Shedden, Will Styler & Daan Wissing. 2018. Plosive voicing in afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics* 66. 185–216. https://doi.org/10.1016/j.wocn.2017.09.009.

Fowler, C. A. 1986. An event approach to the study of speech perception from a direct–realist perspective. *Journal of Phonetics* 14(1). 3–28. https://doi.org/10.1016/S0095-4470(19)30607-2.

Francis, A. L., K. Baldwin & H. C. Nusbaum. 2000. Effects of training on attention to acoustic cues. *Perception & Psychophysics* 62(8). 1668–1680. https://doi.org/10.3758/BF03212164.

Harrington, Jonathan, Felicitas Kleber & Ulrich Reubold. 2012. The production and perception of coarticulation in two types of sound change in progress. In Susanne Fuchs, Melanie

Weirich, Daniel Pape & Pascal Perrier (eds.), *Speech planning and dynamics*, 39–62. Frankfurt: Peter Lang.

Hauser, Irene. 2023. Differential cue weighting in mandarin sibilant production. *Language and Speech* 66(4). 1056–1090. https://doi.org/10.1177/00238309231152495. (Original work published 2023).

Hazan, V. & S. Rosen. 1991. Individual variability in the perception of cues to place contrasts in initial stops. *Perception & Psychophysics* 49(2). 187–200. https://doi.org/10.3758/BF03205038.

Holt, L. L. & A. J. Lotto. 2006. Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America* 119(5). 3059–3071. https://doi.org/10.1121/1.2188377.

Howell, Peter. 1993. Cue trading in the production and perception of vowel stress. *The Journal of the Acoustical Society of America* 94(4). 2063–2073. https://doi.org/10.1121/1.407479.

Izenman, Alan J. 2008. Linear discriminant analysis. In Alan J. Izenman (ed.), *Modern multivariate statistical techniques: Regression, classification, and manifold learning*, 237–280. Springer. https://doi.org/10.1007/978-0-387-78189-1$_8$.

Kang, Yoonjung, Tae-Jin Yoon & Sungwoo Han. 2015. Frequency effects on the vowel length contrast merger in seoul korean. *Laboratory Phonology* 6(3–4). 469–503. https://doi.org/10.1515/lp-2015-0014.

Kleber, Felicitas. 2020. Complementary length in vowelconsonant sequences: Acoustic and perceptual evidence for a sound change in progress in bavarian german. *Journal of the International Phonetic Association* 50(1). 1–22. https://doi.org/10.1017/S0025100317000238.

Kuang, Jianjing & Aletheia Cui. 2018. Relative cue weighting in production and perception of an ongoing sound change in southern yi. *Journal of Phonetics* 71. 194–214. https://doi.org/https://doi.org/10.1016/j.wocn.2018.09.002.

Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017. lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13). 1–26. https://doi.org/10.18637/jss.v082.i13.

Lee, Hyunjung & Allard Jongman. 2018. Effects of sound change on the weighting of acoustic cues to the three-way laryngeal stop contrast in korean: Diachronic and dialectal comparisons. *Language and Speech* 62(3). 509–530. https://doi.org/10.1177/0023830918786305. (Original work published 2019).

Lehnert-LeHouillier, Heidi. 2010. A cross-linguistic investigation of cues to vowel length perception. *Journal of Phonetics* 38(3). 472–482. https://doi.org/10.1016/j.wocn.2010.05.003.

Liaw, Andy & Matthew Wiener. 2002. Classification and regression by randomforest. *R News* 2(3). 18–22. `https://CRAN.R-project.org/doc/Rnews/`.

Liberman, A. M. & I. G. Mattingly. 1985. The motor theory of speech perception revised. *Cognition* 21(1). 1–36. https://doi.org/10.1016/0010-0277(85)90021-6.

Morrison, G. S. 2005. An appropriate metric for cue weighting in l2 speech perception: Response to escudero and boersma. *Studies in Second Language Acquisition* 27(4). 597–606.

Pierrehumbert, J. B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In J. L. Bybee & P. J. Hopper (eds.), *Typological studies in language*, vol. 45, 137–157. John Benjamins Publishing Company. https://doi.org/10.1075/tsl.45.08pie.

Repp, B. H. 1982. Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin* 92(1). 81–110. https://doi.org/10.1037/0033-2909.92.1.81.

Ripley, Brian, Bill Venables, Douglas M. Bates, Kurt Hornik, Albrecht Gebhardt, David Firth & Michael B. Ripley. 2013. *Package 'mass'*. https://cran.r-project.org/web/packages/MASS/. CRAN Repository.

Schertz, J., T. Cho, A. Lotto & N. Warner. 2015. Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics* 52. 183–204. https://doi.org/10.1016/j.wocn.2015.07.003.

Schertz, J. & E. J. Clare. 2020. Phonetic cue weighting in perception and production. *WIREs Cognitive Science* 11(2). https://doi.org/10.1002/wcs.1521.

Shultz, A. A., A. L. Francis & F. Llanos. 2012. Differential cue weighting in perception and production of consonant voicing. *The Journal of the Acoustical Society of America* 132(2). EL95–EL101. https://doi.org/10.1121/1.4736711.

Sonderegger, Morgan & Mrton Sskuthy. 2025. Advancements of phonetics in the 21st century: Quantitative data analysis. https://doi.org/10.31234/osf.io/mc6a9$_v$2.

Weiss, Rudolf. 1974. Relationship of vowel length and quality in the perception of german vowels. *Linguistics* 12(123). 59–70. https://doi.org/10.1515/ling.1974.12.123.59.

Yin, Y. 2024. *Production-perception relationship in bilingual-context sound change: The case of vowel length contrast in longan zhuang*. Beijing: Peking University MA thesis.

Yu, A. C. L. 2019. On the nature of the perception-production link: Individual variability in english sibilant-vowel coarticulation. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 10(1). 1–29. https://doi.org/10.5334/labphon.97.

Zellou, G. 2017. Individual differences in the production of nasal coarticulation and perceptual compensation. *Journal of Phonetics* 61. 13–29. https://doi.org/10.1016/j.wocn.2016.12.002.

Zhang, J., M. Liang & J. Ouyang. 1999. *Zhuangyu fangyan yanjiu [studies on zhuang dialects]*. Sichuan minzu chubanshe.