Journal of Phonetics 71 (2018) 194–214

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Research Article

Relative cue weighting in production and perception of an ongoing sound change in Southern Yi

Jianjing Kuang *, Aletheia Cui

Department of Linguistics, Suite 300, C-3401 Walnut Street, Philadelphia, PA 19104, United States

ARTICLE INFO

Article history:
Received 11 October 2017
Received in revised form 7 September 2018
Accepted 8 September 2018
Available online 3 October 2018

Keywords:
Sound change
Cue weighting
Production and perception

ABSTRACT

Multiple co-varying cues for a phonological contrast are often introduced by coarticulation, and sound change occurs when their relative weighting shifts. The central issues for this kind of sound change include how cue weighting shifts over time in both production and perception and what the mapping is between production and perception during this process. This study aims to provide insights to these questions by examining an ongoing change in the tense vs. lax register contrast in Southern Yi. Production and perception experiments were conducted with the same group of speakers to evaluate the relative importance of the source cue (i.e., phonation) and its coarticulated cues (i.e., vowel formants and f0) for this contrast. While speakers of all age groups still maintain the register contrast, our results show that formant differences are overtaking phonation as the primary cues. This sound change is more advanced in non-high vowels than high vowels in both perception and production. Moreover, production and perception are misaligned; in both cases, the shift to formant values occurs first in perception, with production lagging behind. These findings illustrate the nuanced progression of sound change and a better understanding of the role of production and perception in the initiation of a sound change.

1. Introduction

Phonological contrasts are usually realized with multiple co-varying cues, and sound change can occur when the primary cue of the given contrast shifts. Tonogenesis is a well-known case for this type of sound change – while the original lexical contrasts are maintained, pitch overtakes other cues such as voicing as the primary cue. An important theoretical question about this type of sound change is how and when the primary cue shifts in production and perception. In particular, do speakers or listeners lead the shift? What is the interaction between production and perception during the cue shifting in progress? The goal of this study is to provide insights to these questions by examining an ongoing cue-shifting change in Southern Yi, where vowel quality is overtaking phonation as the primary cue for its register contrast. This understudied sound change is analogous to tonogenesis in many ways, and thus can provide us with an important case study to better understand the initiation stage of this type of sound change *in vivo*.

1.1. Cue shifting and sound change

1.1.1. The role of co-varying cues in sound change

Speech signals are highly redundant. In speech production, an articulatory target is often achieved by the coordination of multiple articulators. As a result, this process, known as coarticulation,[1] is one of the major causes for a given phonological contrast to have multiple co-varying acoustic cues. For example, Lisker (1986) noted that the voiced vs. voiceless contrast of English obstruents involves at least 16 acoustic cues, including the intensity of the glottal signal, the duration of the vowel, the duration of the first formant transition, F1 offset frequency, voice onset time (VOT), f0 contour, and so on. While listeners take advantage of multiple cues to ensure the success of perceiving the intended linguistic targets (e.g., Brunelle, 2012; Kingston, Diehl, Kirk, & Castleman, 2008; Kuang, 2013; Toscano & McMurray, 2010), the co-varying cues differ in their contribution, or weights, to a phonological contrast, as listeners' attention is selective. For example, among the large set of co-varying

* Corresponding author.
E-mail address: kuangj@ling.upenn.edu (J. Kuang).

[1] Coarticulation is also commonly used to refer to the contextual influence from neighboring segments.

https://doi.org/10.1016/j.wocn.2018.09.002
0095-4470/© 2018 The Authors. Published by Elsevier Ltd.

cues, voice onset time (VOT) has been found to be the most reliable acoustic cue for the English voicing contrast in the onset position (Lisker & Abramson, 1964, 1970; Schertz, Cho, Lotto, & Warner, 2015; Francis, Kaganovich, & Driscoll-Huber, 2008; Davidson, 2016; Kong & Edwards, 2016; Nelson & Wedel, 2017).

Nonetheless, secondary cues can provide opportunities for sound change, which may occur when the relative weighting of the cues shifts. One of the best-known instances of this kind of sound change is tonogenesis. It has been well-established that vowels following a voiceless onset tend to have higher f0 (e.g., Hombert, Ohala, & Ewan, 1979; Kingston, 2005; Maddieson, 1984; Ohala, 1973; Thurgood, 2002), and this f0 perturbation effect is fairly common among languages (see Hombert et al., 1979 for Yoruba, Xu & Xu, 2003 for Mandarin, Haggard, Ambler, & Callow, 1970 for English, and Jun, 1996 for Korean and French). Notably, in some languages, such as Vietnamese (Thurgood, 2002), this synchronic variation in f0 eventually developed into phonemic contrasts, i.e., tones. In this kind of change, while a phonological contrast is preserved, one cue overtakes another as primary. Before tonogenesis, the primary contrastive cue is the voicing of the onset consonant. After tonogenesis, the primary cue for the contrast has shifted to the pitch differences on the vowels, and the voicing contrast on the consonants is often completely lost (Hyman, 1976; Hombert et al., 1979; Kingston, 2005; Kirby, 2013; Thurgood, 2002). An important question is how a coarticulated cue rises in significance and becomes the primary cue.

According to Hyman (1976), tonogenesis involves three steps: (1) voiced and voiceless consonants determine the f0 perturbations on following vowel as the result of intrinsic coarticulation, (2) f0 perturbations are exaggerated and become a perceptual cue (i.e., phonologized), and (3) distinct tones develop and the consonant voicing distinction is lost (i.e., phonemicization). In this proposal, phonologization in step 2 is crucial for synchronic variation to turn into sound change, during which an acoustic cue (e.g., f0) becomes a significant but secondary contributor in both the production and the perception of voicing. However, this step by itself is not enough to trigger sound change (e.g., tonogenesis), because it is natural for languages to have multiple stable secondary cues (Wright, 2004) as enhancement cues for the target contrast (Stevens & Keyser, 2010). For example, f0 (Haggard, Kleber, & Reubold, 1970), F1 formant transition (Liberman, Delattre, & Cooper, 1958), vowel duration (Summerfield, 1981), aspiration amplitude (Repp, 1979), and burst spectrum (Chodroff & Wilson, 2014) all significantly contribute to the voicing contrast in both production and perception for English speakers, but English is not undergoing tonogenesis. Actual tonogenesis occurs only when the contrast is reanalyzed and the primary cue of the contrast shifts from VOT to f0. In other words, there must be additional intermediate steps between step 2 and step 3, where the novel phonologized secondary cue rises in significance and becomes the primary cue in production and perception. The question is, then, how is the change of the primary cue implemented in production and perception?

1.1.2. The mapping between perception and production during sound change

In theory, there are three possibilities for the time course of cue shifting:

(1) the primary cue shifts in production and perception at the same time,
(2) listeners first shift their attention to a new cue in perception, and then in turn rely on this cue to mark a phonological contrast in production, and
(3) cue shifting starts in production, and listeners subsequently become attuned to the changes in perception.

Possibility (1) assumes parity between production and perception during sound change. Generally speaking, there should be parity between production and perception because gestures can usually be recovered from the speech signal (Fowler, 2005; Fowler & Smith, 1986). Both possibility (2) and possibility (3) assume that there can be misalignment between production and perception, i.e., speakers rely on different primary cues in production and perception.

Several models of sound change support the scenario in possibility (2). Ohala's model of sound change (Ohala, 1981, 1993) and the extensions thereof (e.g., Solé, 2014) proposed that the driving force of sound change is the unintentional error on the part of the listener. In these proposals, sound change occurs when the listener fails to compensate for the effects of contextual coarticulation or when they attribute the coarticulated effects to a wrong source. While also recognizing that listeners are the driving force of sound change, Beddor (2009) suggested that this kind of parsing is not a mistake on the listener's part. Rather, listeners actively attend to any relevant cues because multiple grammars are consistent with the input. Some listeners simply place more weight on the effect than on the source of the coarticulation. Beddor (2012) also hypothesized that variation in interpreting the cues is earlier than the actual sound change, which is a process that results in the emergence of the innovative cues. Altogether, these theories predict that the reanalysis of the contrastive cues (i.e., the shift of the primary cue) is manifested in perception before production. It should be noted that, these theories also assume that before perceptual reanalysis, the raw materials for cue shifting, such as reliable secondary cues, are already grounded in production.

While possibility (2) states that perception plays an active role in reanalysis, possibility (3) suggests that perception plays a rather passive role, and that cue shifting is largely driven by changes in production. In particular, it has been proposed (e.g., Abramson, 2004; Kang, 2014; Kirby, 2013) that a secondary cue is likely to take over as the primary cue when the original primary cue is neutralizing and merging – that is, to "save" the lexical contrasts, speakers may emphasize a secondary cue to compensate for the loss of the primary cue. Under this proposal, the shifting of the primary cue might happen in production first. Another reasoning (e.g., Janson, 1983) is based on the fact that speakers who utilize different cue weights in production (e.g., old vs. young speakers) can usually perfectly understand each other. In order to maintain the mutual intelligibility in communication, cue shifting in perception might be slower than in production.

1.1.3. Previous work on the production-perception mapping

To better understand the nuanced progression of how cue shifting is initiated in production and perception, it is necessary to validate the above-mentioned hypotheses with empirical data. In order to do so, it is important to conduct comparisons between perception and production from the same group of speakers, ideally based on a case of sound change that is still at the initiating stage. These kinds of empirical studies are still relatively rare, and mixed findings have been reported in the limited literature.

Coetzee, Beddor, Shedden, Styler, and Wissing (2018) studied the relative cue weighting of f0 and prevoicing in the production and perception of plosive voicing in Afrikaans from the same group of speakers. Consistent with the hypothesis of tonogenesis, while all speakers used f0 in their production and perception, older speakers were more likely than younger speakers to rely on prevoicing in production and perception. The different cue preferences between old and young speakers are generally comparable to the production and perception results[2] of tonogenesis in Korean (e.g., Kang & Guion, 2008; Kang, 2009, 2014; Silva, 2006; Wright, 2007). However, there is also a notable difference between these two languages: While cue trading in production (i.e., the simultaneous increase in f0 differences and decrease in VOT differences) was reported in the Korean case (Bang, Sonderegger, Kang, Clayards, & Yoon, 2018; Kang, 2014), the f0 differences were similar between the young and old speakers in Afrikaans (Coetzee et al., 2018). In terms of the relationship between production and perception, the Afrikaans study found that individual speakers generally had aligned production and perception (i.e., speakers relied on the same primary cue in both production and perception), but when there were misalignments, production was in the lead. In other words, even though some of the young speakers no longer produced VOT differences, they were still sensitive to this cue perceptually. A similar trend of production-perception mapping was reported by Pinget (2015) on the case of Dutch devoicing. Overall, at the completion stage of a sound change, perception seems to lag behind.

However, there are also cases in which perception is in the lead. Harrington and colleagues (e.g., Harrington, 2012; Harrington, Kleber, & Reubold, 2008; Kleber, Harrington, & Reubold, 2012) analyzed the production and perception of /u/- and /ʊ/-fronting by the same older and younger speakers of Standard Southern British English. These studies found that on the community level, the link between production and perception is unstable during sound change, and category boundary shift for /u/ entered perception before production. What is striking about this study is that for /ʊ/, there was a similar shift in the perceptual boundary for younger speakers even though no shift was observed in their production. This means that perceptual cue shifting in the direction of the change can be initiated ahead of similar shifts in production.

These empirical studies bring out a few more important questions: (1) Do production and perception have the same relationship (either production in the lead or perception in the lead) throughout the process of sound change? Is it possible that production and perception have different relationships in an earlier stage vs. a later stage, so that the discrepancy reported in the literature is the result of different stages of sound change? (2) Do production and perception have the same relationship across different languages, different types of sound change (e.g., merger, cue shifting, or boundary shifting), and different types of coarticulation? All in all, to understand the relationship between production and perception during a sound change in progress, more empirical studies from different types of sound change should be conducted. The goal of our study is to shed light on the relationship between production and perception during a sound change by looking into an ongoing sound change in Southern Yi.

1.2. The register contrast in Southern Yi

The Yi language belongs to the Loloish branch of the Tibeto-Burman language family. It has a register contrast in its vowel system that is typically realized as distinct phonation – broadly defined as different laryngeal configurations such as different extents of glottal constriction. This contrast is commonly termed the tense vs. lax contrast. Southern Yi is one of the several dialects of the Yi language, and it is primarily spoken in Yunnan, China. Southern Yi contrasts in three tones: low (21), mid (33), and high (55) and has a seven-vowel system: /i, ɛ, a, z̩, ə, o, u/. The register contrast in Yi has been well documented (Southern Yi in Kuang, 2011a, Kuang & Keating, 2014, and other Yi dialects and languages in Maddieson & Ladefoged, 1985, Edmondson & Esling, 2006, Kuang, 2011b, Kuang & Keating, 2014). While the tense vs. lax register contrast co-occurs with all the vowels, it only co-occurs with the low (21) and mid (33) tones and is not found with the high (55) tone. Therefore, with the combination of tone and register dimensions, it is possible for a given syllable to have a five-way contrast. For instance, the syllable "be" (phonetically /bɛ/, we will use broad transcription /be/ hereafter) can have the following register and tone combinations (the tense register is conventionally denoted by an underline): /be_21/ "to entangle", /be21/ "to drop (something)", /be_33/ "to shoot (an arrow)", /be33/ "to argue", and /be55/ "jug".

1.2.1. The acoustic cues for the register contrast

The acoustic and articulatory properties of the tense and lax contrast have been well established in previous studies (e.g., Kuang, 2011a; Kuang & Keating, 2014). The tense phonation has relatively greater glottal constriction, as evidenced by a greater Contact Quotient in the electroglottographic (EGG) signal. Acoustically, the greater glottal constriction results in a less prominent first harmonic (H1) in the spectrum. To normalize the recording effects, the prominence of H1 is usually measured as the relative amplitude differences between the first harmonic and some higher frequency components in the spectrum, such as H1-H2 (relative to the second harmonic), H1-A1, H1-A2, and H1-A3 (relative to the highest amplitude around the first three formants: namely, A1, A2, and A3). This set of spectral measures has been found to be significant in distinguishing the two registers in Southern Yi (Kuang, 2011a). As shown in Fig. 1, the tense phonation exhibits smaller H1-H2, H1-A1, H1-A2 and H1-A3, while the lax phonation shows greater values for these spectral measures. In addition, the tense phonation is also generally more periodic and less noisy, as

[2] However, these production and perception experiments were not conducted with the same group of speakers.

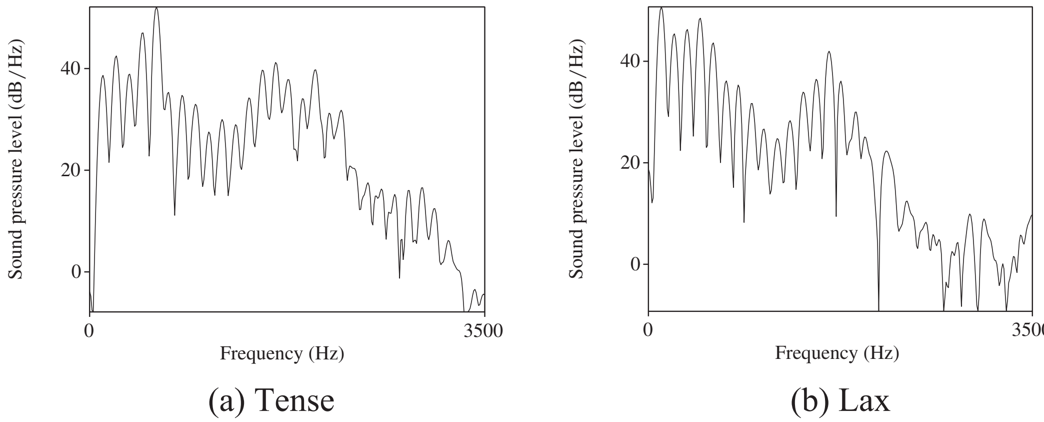


Fig. 1. Spectral differences for the tense /be_/ and the lax /be/ in Southern Yi.

shown by greater values of Cepstral Peak Prominence (CPP) and Harmonic-to-Noise Ratio (Kuang, 2011a).

In addition to distinctive glottal constrictions and their spectral correlates, the register contrast in Yi also typically exhibits differences in F1, F2, or f0 due to the coarticulation between multiple glottal and supraglottal articulators (Edmondson & Esling, 2006; Kuang, 2011a). It has been well-established that stiffness in the vocal folds often increases the rate of vibration, which leads to a higher f0 (e.g., Löfqvist, Baer, McGarr, & Seider Story, 1989; Titze, 1990; Zhang, 2016). Indeed, significantly higher f0 has been observed for tense vowels in languages related to Southern Yi, such as Bo (Kuang, 2011b) and Eastern Yi (Maddieson & Ladefoged, 1985). Moreover, supraglottal settings are usually involved in the phonation contrast. A laryngoscope study of Northern Yi (Edmondson, Esling, Harris, Shaoni, & Ziwo, 2001) found that the production of the tense phonation involves retracting the tongue root and raising the larynx, so the tense vowels have relatively lower and more back vowel quality. As a result, the F1 of tense vowels tends to be higher, and the F2 tends to be lower. These F1 and F2 effects were found for Southern Yi as well (Kuang, 2011a, p. 59)[3], and similar supraglottal articulations are likely to be involved.

1.2.2. Coarticulated cues and sound change in Southern Yi

Notably, different coarticulation patterns for the phonation contrast both across different languages and within the same language can determine the possible paths of sound change. For example, Kuang (2011b) found that the tense and lax registers in Southern Yi are produced with consistent F1 differences but little f0 differences, whereas in Bo, a related language, f0 but not F1 plays a significant role for the tense vs. lax contrast. This variation suggests different directions of potential sound change for these two languages: Vowel-splitting is more likely for Southern Yi, and tone-splitting is more likely for Bo. Furthermore, within Southern Yi, the coarticulatory effect of tongue root retraction applies to different extents depending on vowel height. Kuang (2011a) observed that F1 differences of the register contrast are more salient for the non-high vowels (i.e., /ɛ/, /a/) than for high vowels (i.e., /i/, /u/).

Kuang (2011a) also found speaker variation in terms of the production of the register contrast: For some younger speakers, the non-high vowel pairs were produced with only vowel quality differences and no phonation differences, but the high vowel pairs were mostly distinguished by phonation differences. Therefore, it is likely that a sound change is in progress in this Southern Yi dialect, especially since a few other Yi dialects have completed this change and use vowel quality instead of phonation for the register contrast. For example, Kuang (2011a) noted that the corresponding words of the /ɛ_/ vs. /ɛ/ contrast in a closely related Yi dialect have become /æ/ vs. /ɛ/ (lower vowel /æ/ corresponding to tense /ɛ_/). Moreover, in Northern Yi, it has been documented that the phonation contrast only occurs with high vowels, and the register contrast for non-high vowels is entirely realized as vowel quality differences (Edmondson et al., 2001). Therefore, a vowel split originating from a phonation-based register contrast is a natural sound change that commonly occurs among the Yi dialects, and we are very fortunate to be able to observe the initiation stage of this sound change *in vivo* in Southern Yi.

1.3. Summary of the research questions

This sound change in Southern Yi is analogous to tonogenesis in many ways since it is another typical case of the cue shifting type of sound change. As we discussed in detail in section 1.1, for tonogenesis, it is pitch that overtakes voicing as the primary cue; for the sound change in Southern Yi, it is vowel quality that overtakes phonation as the primary cue. Therefore, the ongoing sound change in Southern Yi provides us with an important case study to answer questions about the cue shifting type of sound change.

In our current study, by conducting cue weighting experiments in perception and production for high vowels and non-high vowels with the same group of Southern Yi speakers, we can address the following research questions:

1. Are production and perception aligned during the cue shifting type of sound change? If not, does production or perception lead the change?
2. Do production and perception have the same mapping relationship for older and younger speakers, since younger speakers are likely to be more advanced in this sound change?

[3] F1 effects are consistent across all vowels, but F2 effects vary for some vowels.

3. Is it true that non-high vowels are more advanced in this sound change than high vowels? If so, do production and perception have the same mapping relationship for high vowels and non-high vowels?

2. Experiments

2.1. Participants

All the experiments in this study were conducted in the Xinping village in Yunnan, China. All experiments were conducted with the same 41 native speakers of Southern Yi from this village. There were 17 female speakers (age range 26–70, mean 47) and 26 male speakers (age range 30–71, mean 45). The participants were divided into three age groups: younger than 40 (10 participants), 40–50 (17 participants), and older than 50 (14 participants). All speakers but one also spoke a dialect of Southwestern Mandarin.

2.2. Perception experiment 1: natural stimuli

This experiment tested the participant's perception of the phonation contrast in naturally produced speech in high vowels and non-high vowels. The goal of this experiment was to evaluate whether the register contrast was maintained by all age groups in perception. The results provided a baseline for comparison with further experiments in cue weighting between the different groups.

2.2.1. Stimuli

This experiment used the naturally produced minimal sets of /be/ (be̲21, be21, be̲33, be33, and be55) and /bu/ (bu̲21, bu21, bu̲33, bu33, and bu55) from a previous production experiment (Kuang, 2011a), shown in Table 1. These two sets were selected because they were the minimal sets native speakers found most familiar; all the words are relatively frequent, and all speakers were able to recognize and produce these minimal sets without difficulty. The stimuli were taken from the production of 3 female speakers and 3 male speakers, all of them in their 40 s (the middle age group in this study). Each word was produced in isolation. There were 120 tokens in a total: 5 contrasts per set × 2 minimal sets × 6 speakers × 2 repetitions.

2.2.2. Procedures

Before the perception experiment, the two minimal sets were first elicited from the participants in production. They were asked to produce the minimal sets in a carrier sentence and then form sentences using the target words. This procedure was intended to ensure that all the test tokens were fully familiarized and activated for the participants so that they would have a relatively equal expectation for all test tokens, therefore minimizing any potential lexical frequency bias in the responses. The experiment was administered in a quiet room on a laptop through Praat. In the identification task, each stimulus was presented in isolation, and instructions were presented on the screen in Mandarin Chinese. There were five possible choices for each stimulus, each corresponding to a word in the minimal set. The choices were presented in Chinese characters because we found this to be a more effective way of eliciting responses than using either pictures or Yi characters based on our experience with the participants. The order of the stimuli was randomized for each session. The participants were asked to select the choice that best corresponded to the stimulus. The experiment was self-paced. Listeners were able to replay each stimulus as many times as necessary before responding, and they were allowed to go back and re-listen to previous stimuli. After making a choice, they could go on to the next stimulus by clicking the "next" button.

Table 1
Minimal sets of /be/ and /bu/ words used for the experiments.

	21T	21L	33T	33L	55L
/be/	to entangle	to drop	to shoot	to argue	jug
/bu/	mold	to carry	to be full	worm	to cry

2.2.3. Results

The identification rates of tone × register categories are presented in Figs. 2 and 3. It should be noted that identifying tone × register categories across multiple speakers is a difficult task, requiring effective speaker normalization. Nonetheless, the listeners were able to reliably identify the phonological categories. As shown in Figs. 2 and 3, the identification rates are well above the 20% chance level for all phonological categories except for /be55/. Post-test interviews with the subjects revealed that cause for the low accuracy rates for /be55/ was likely lexical rather than phonetic or phonological. The lexical item typically pronounced as /be21/

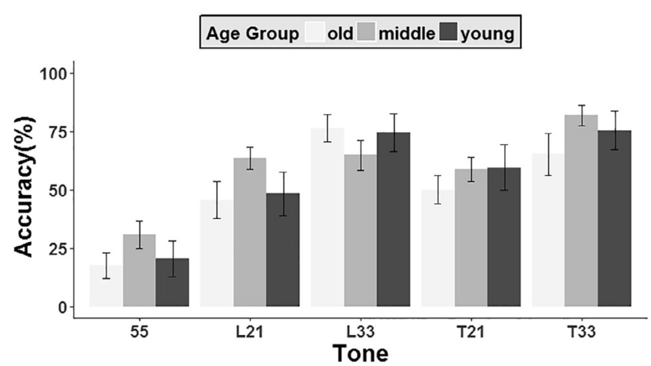


Fig. 2. Identification rates for /be/: the x-axis lists the five tone × register categories; age groups are indicated by color.

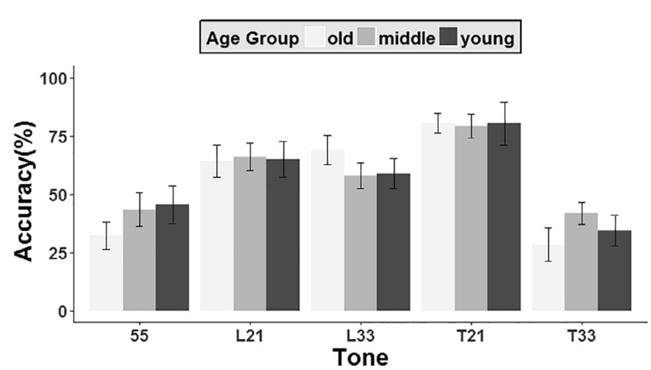


Fig. 3. Identification rates for /bu/: the x-axis lists the five tone × register categories; age groups are indicated by color.

has a secondary pronunciation of /be55/, leading a number of subjects to choose the character for /be21/ upon hearing /be55/. As for other phonological categories, it is not surprising that some categories have relatively poorer identification rates. For example, Lax 55, Tense 33 and Lax 33 are phonetically similar in terms of phonation and pitch, and they are thus more likely to be confused with each other. Most importantly, all age groups experienced similar levels of difficulty in the identification of each of the categories. To test whether age groups had any effect on the perception accuracy, a logistic mixed-effects model was fitted for each syllable type (i.e., /be/ and /bu/) using the *glmer* function from the *lme4* package (Bates, Maechler, & Bolker, 2013) in R (R Core Team, 2017), with *p*-values estimated using the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2016) in R. The dependent variable of the model was accuracy (i.e., whether the word was accurately identified); age_group was included as the fixed factor, and stimulus item and speaker were included as random intercepts. To obtain comparisons between every two age groups, the same model was run twice with two different reference levels. As summarized in Table 2, the results confirmed that there were no age differences in the overall accuracy rates.

To further test whether listeners from all age groups were able to reliably distinguish the tense register from lax, another mixed-effects logistic regression model was constructed for each syllable type, with responses of tense (combining T21 and T33) or lax (combing L21, L33, and L55) as the binary dependent variable, register (Tense vs. Lax) and age group (young, middle, old) and their interaction as the fixed factors, and listener as the random intercepts. Since both of the predictors are categorical and contain multiple levels, a likelihood ratio test using the *anova* function should be performed to evaluate the overall significance of each factor and their interactions (Baayen, 2008, pp. 199–200). Because *anova* is not applicable to *glmer* models, we used the *PBmodcomp* function from the *pbkrtest* package (Halekoh & Højsgaard, 2014) to conduct the likelihood ratio test. The results are summarized in Tables 3 and 4. For both /be/ and /bu/, register has a significant main effect on the classification of the tense vs. lax registers. Also, there was no significant interaction between

Table 2
Summary of the mixed-effects logistic regression models: age effects on perception accuracy.

Syllable: /be/	young	middle
middle	$\beta = -0.172$, $p > 0.1$	
old	$\beta = -0.113$, $p > 0.1$	$\beta = 0.172$, $p > 0.1$
Syllable: /bu/	**young**	**middle**
middle	$\beta = 0.043$, $p > 0.1$	
old	$\beta = 0.015$, $p > 0.1$	$\beta = 0.056$, $p > 0.1$

Table 3
Summary of the mixed-effects logistic regression models: age effects on register classification (syllable /bu/. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$).

	LRT	p-value
register	658.47	0.02^{*}
age_group	1.52	0.41
register:age_group	3.05	0.18

Table 4
Summary of the mixed-effects logistic regression models: age effects on register classification (syllable /be/. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$).

	LRT	p-value
register	1407.61	0.02^{*}
age_group	1.07	0.49
register:age_group	1.59	0.57

age groups and register, further confirming that all age groups were able to reliably classify the tense vs. lax registers.

In sum, experiment 1 confirms that the tense and lax contrast is maintained by all age groups. We can therefore move on and ask: While all age groups are able to perceive the register contrast, do they rely on the same sets of cues to do so? Cue weighting experiments in production and perception were conducted in order to answer this question.

2.3. Experiment 2: cue weighting in production

A production experiment was conducted to assess the relative importance of cues the speakers used in producing the tense vs. lax contrast and whether speakers of different ages differed in their use of these cues. We predict that if Southern Yi is undergoing sound change, there should be cue weighting differences between older speakers and younger speakers in the direction of the change. Specifically, younger speakers are expected to use vowel quality cues more (e.g., F1 and/or F2) to produce the register contrast, while older speakers are expected to use phonation cues more.

2.3.1. Methods

Recordings were made in a quiet room in the village. The speech materials comprised minimal sets of monosyllabic words /be/ and /bu/ with all possible register and tonal combinations, the same as the sets used in the previous perception experiment. To avoid the influence of sentence-final prosody (which usually introduces creaky voice), all target words were produced in the frame: [ŋo33 __ e55 lɤ33 ɣw33] "I say the word ___." To obtain careful speech which was desired for reliable phonation measures, words were elicited in minimal sets, but the speakers were not told that these words would have different pronunciations. Each utterance was repeated three times. The audio signal was recorded through a Shure WH-30 head-worn unidirectional microphone at a sampling rate of 22,050 Hz. The vowel portion of each target word was extracted, and acoustic measurements that are relevant to the register contrast, including phonation cues, vowel quality cues (F1, F2), and pitch (f0), were taken automatically using VoiceSauce over the entire vowel (Shue et al., 2011). The phonation cues included $H1^*$, $H1^*$-$H2^*$, $H1^*$-$A1^*$, $H1^*$-$A2^*$, $H1^*$-$A3^*$ and CPP. To allow for cross-vowel comparisons in voice quality, the spectral tilt measures were corrected for vowel formants, and the corrected measures are marked with asterisks. Following previous studies (Kuang, 2011a; Keating et al., 2011), mean values of the acoustic measures were used in the statistical analysis.

2.3.2. Results

The raw data values were within-speaker z-score normalized (Lobanov, 1971) based on each speaker's mean values before statistical analysis. Since phonation is correlated with multiple measures, a principal component analysis was run on the phonation measures (i.e., H1*, H1*-H2*, H1*-A1*, CPP, H1*-A2*, H1*-A3*) that were found to successfully distinguish the register contrast in Southern Yi (Kuang, 2011a). The first principal component (PC1 hereafter) accounted for 53% of the variance, and the second principal component (PC2) accounted for 21% of the variance. PC1 was highly correlated with H1*-A1* ($r = 0.93$), H1*-A2* ($r = 0.\ 9$), and H1*-A3*($r = 0.91$), and PC2 was highly correlated with H1*-H2* ($r = 0.8$) and H1* ($r = 0.66$). Overall, the PCA results confirm the previous claims in Kuang (2011a) that the strength of the first harmonic (H1) relative to higher frequency components in the spectrum is the most important acoustic correlate for the phonation differences in the tense vs. lax register contrast in Southern Yi. Therefore, PC1 was used to represent the phonation cue in the regression models described below. The F1 × F2 space as well as the phonation PC1 values are plotted in Figs. 4 and 5.

As shown in Fig. 4, the overall F1 × F2 range is fairly similar between the three age groups. Mean F1 differences between the tense and lax /be/ were calculated for the three age groups, showing that young speakers produce slightly larger F1 difference (131.9 Hz) than older speakers (121.4 Hz for the middle group, and 122.4 Hz for the old group), but a 10 Hz difference is not likely to be perceptible. Similar calculations were done for F2 as well, and there is also no evidence that F2 differences between the tense and lax /be/ is larger for the young group (old: 143 Hz, middle: 102 Hz, young: 114 Hz). Overall, since younger speakers of Southern Yi did not produce more distinct F1 and F2 than older speakers, there is no evidence that formant cues are enhanced by younger speakers when compared to older speakers. However, when we turn to the phonation cue, as shown in Fig. 5a, there is a clear trend that phonation differences between the two registers are reduced for the young speakers when they produce syllable /be/. By contrast, for /bu/, as shown in Figs. 4 and 5b, both the vowel space and the phonation distinction are fairly similar across the three age groups.

These observations were confirmed by statistical analysis. To estimate the relative importance of F1, F2, f0, and phonation in the production of the register contrast, and whether the relative contribution of these acoustic cues is significantly modulated by age groups, a mixed-effects logistic regression model was built for each syllable type (i.e., /be/ or /bu/), using the *glmer* function from the *lme4* package (Bates et al., 2013) in R (R Core Team, 2017), with *p*-values estimated using the *lmerTest* package (Kuznetsova et al., 2016) in R. The dependent variable was the binary tense vs. lax register contrast,

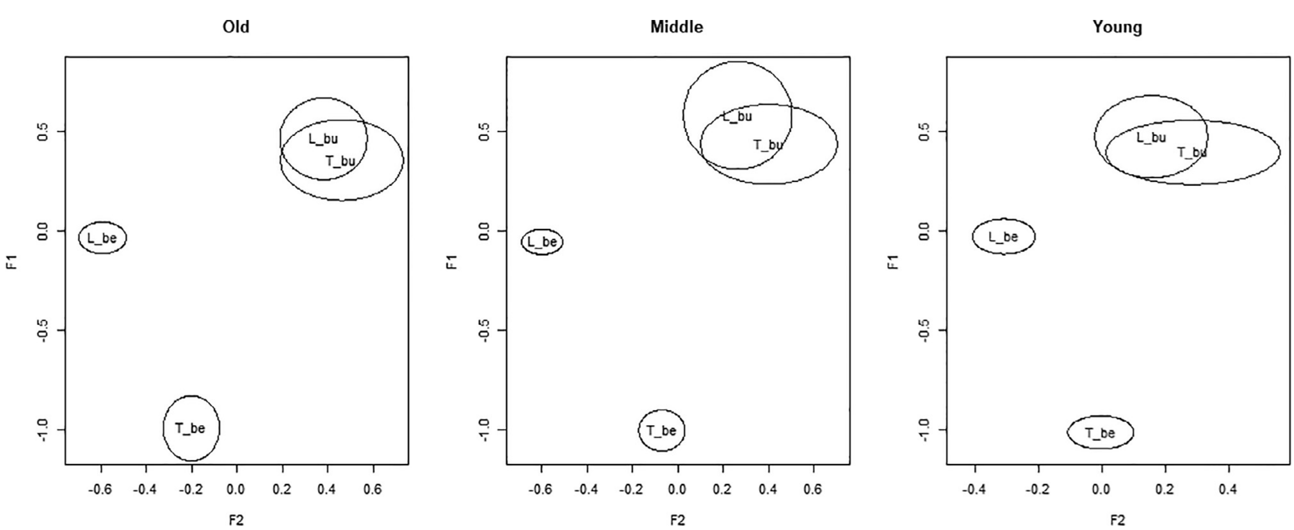
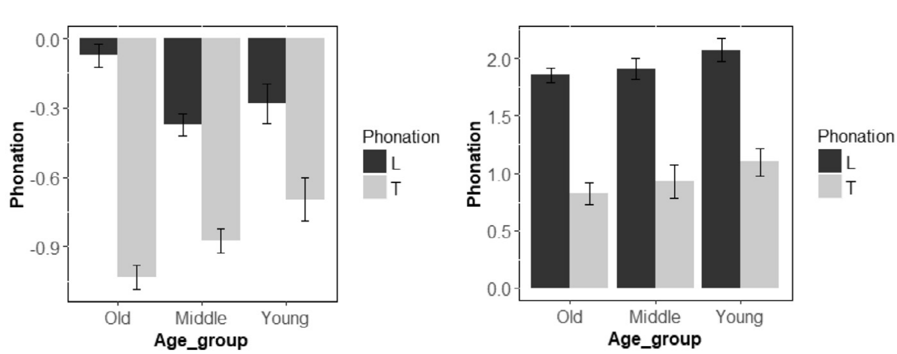


Fig. 4. Within-speaker normalized F1 × F2 spaces for tense vs. lax register contrast across syllable types and age groups.



(a) /be/ (b) /bu/

Fig. 5. Within-speaker normalized phonation distinction (PC1 of multiple phonation measures) for the register contrast for the three age groups. (a) Syllable /be/; (b) syllable /bu/.

and the fixed factors were F1, F2, f0, phonation PC1, and interactions of each acoustic cue with age groups. Speakers were included as the random intercepts. The models used the default dummy coding system, and the old speaker group was set as the reference level for the age_group factor. The models failed to converge when random slopes were added, so only random intercepts were retained. The results are summarized in Tables 5 and 6.

For /be/, as shown in Table 5, F1, phonation, and F2 contributed significantly to the classification of the register contrast. More importantly, there are significant interactions between phonation and the young group, and F1 and the young group. There is also a marginal interaction between F2 and the young group. In general, this model shows a shift in the relative cue weighting between phonation and formants, especially F1. There are three possible reasons for this shift: (1) Similar phonation distinction, but more enhanced formant differences; (2) similar formant differences, but reduced phonation differences; (3) a trading relationship in which a reduced phonation difference is accompanied by increased formant differences. Based on Figs. 4 and 5a, the cue shift for /be/ is consistent with the second scenario. Therefore, the cue shifting found for /be/, as evidenced by the significant interaction of young group × F1 as well as young group × phonation found in Table 5, is mostly driven by the reduction of the phonation distinction by the young speakers, similar to the case of Afrikaans (Coetzee et al., 2018).

Table 5
Summary of the mixed-effects logistic regression model for the production of /be/ (dummy-coded, age_group reference level = old speakers. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$).

Syllable /be/	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	−2.95	0.46	−6.40	0.00**
F1	1.63	0.51	3.17	0.00**
Age_group(middle)	0.15	0.80	0.19	0.85
Age_group(young)	0.27	0.81	0.34	0.73
Phonation PC1	−1.83	0.32	−5.73	0.00***
F2	−0.87	0.42	−2.05	0.04*
F0	−0.18	0.21	−0.84	0.40
F1:Age_group(middle)	0.59	0.71	0.84	0.40
F1:Age_group(young)	2.57	1.20	2.14	0.03*
Age_group(middle):Phonation PC1	0.28	0.39	0.71	0.48
Age_group(young):Phonation PC1	1.13	0.45	2.52	0.01*
Age_group(middle):F2	−0.05	0.71	−0.07	0.95
Age_group(young):F2	−2.04	1.04	−1.96	0.05.
Age_group(middle):F0	−0.42	0.34	−1.22	0.22
Age_group(young):F0	−0.30	0.39	−0.75	0.45

Table 6
Summary of the mixed-effects logistic regression model for the production of /bu/ (dummy-coded, age_group reference level = old speakers. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$).

Syllable /bu/	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.18	0.29	4.11	0.00**
F1	0.03	0.20	0.14	0.89
Age_group(middle)	0.10	0.48	0.21	0.83
Age_group(young)	0.37	0.54	0.67	0.50
Phonation PC1	−1.17	0.17	−6.86	0.00***
F2	−0.04	0.18	−0.21	0.84
F0	−0.23	0.14	−1.56	0.12
F1:Age_group(middle)	−0.03	0.31	−0.10	0.92
F1:Age_group(young)	−0.20	0.43	−0.46	0.64
Age_group(middle):Phonation PC1	0.11	0.27	0.39	0.70
Age_group(young):Phonation PC1	−0.08	0.30	−0.28	0.78
Age_group(middle):F2	0.33	0.28	1.16	0.25
Age_group(young):F2	0.33	0.36	0.92	0.36
Age_group(middle):F0	0.24	0.23	1.07	0.28
Age_group(young):F0	0.48	0.25	1.93	0.06

By contrast, for /bu/, as shown in Table 6, phonation is the only significant predictor for the register contrast, and there is no significant interaction between age group and any of the acoustic cues. This suggests that speakers from all age groups produced the tense vs. lax register contrast in the same way and confirms the observations made from Figs. 4 and 5b. Overall, this experiment shows that in production, there is cue shifting for /be/, but not for /bu/.

2.4. Experiment 3: cue weighting in perception

Given that the production experiment showed differences between /be/ and /bu/, it is possible that phonation remains the primary cue for /bu/ in perception, but not for /be/. Moreover, since the production experiment also showed age differences, if perception matches production, a second prediction is that there might be age differences in the cue weighting in perception: older speakers may rely more on phonation, and younger speakers may rely more on vowel quality. To test these hypotheses, we needed stimuli that provided different combinations of cues. Therefore, a perceptual identification experiment with resynthesized stimuli was conducted to explore cue weighting in the perception of the register contrast by different age groups as well as the differences between high vowels and non-high vowels.

2.4.1. Stimuli

Synthesized stimuli were created from the naturally produced tokens of a male speaker and a female speaker from a previous production experiment (Kuang, 2011a), and both speakers were in their 40s. For our purposes, we needed speakers who produced the register contrast with distinct phonation. These two particular speakers were chosen because their production of the register contrast was best categorized by listeners in experiment 1. Two different genders were used because, given that female speakers are often leaders of sound change (Labov, 1994), cue weighting might be gender-specific.

Two minimal pairs that contrast in phonation were selected: /be33/ “to shoot (an arrow)” vs. /be_33/ “to argue” and /bu33/ “worm” vs. /bu_33/ “to be full”. Before manipulating the target cues, the natural tokens were neutralized in duration, peak amplitude, VOT, and burst noise. Next, the phonation of the original tokens was maintained, and F1, F2, and f0 of these tokens were modified using Praat in incremental steps (5 F1 × 3 F2 × 3 f0) for each of the tokens. The ranges of the F1 and F2 continua were based on the natural contrast of the original tokens, but the extreme ends were expanded from the natural differences between the tense and lax registers of any age group. The formant values were adjusted using LPC resynthesis in Praat, the sampling rate of 22,050 Hz of the original tokens was maintained. For /bu/, the natural F1 distinction between tense and lax vowels is relatively small, but in this perception experiment, we used the same range of 300 Hz as the /be/ pair so that the stimuli did not contain inherent bias towards any particular age group. The f0 range was chosen in order to cover the natural variation in the speakers’ production. In the resynthesis, the onset of each vowel was set at the specified values for each f0 step. The ending point of the vowel was 5 Hz below the starting point to simulate the slight natural decline in pitch for level tones in natural production. The f0 falls from the starting value to the ending value in a smooth slope.

Overall, the manipulated cues were unlikely to cause the listeners to identify the stimuli voice with any particular age group, and the only chance the listeners might be able to tell the age of the speakers was from the unmanipulated phonation cues. In a follow-up study (Kuang and Cui, 2018), we asked listeners (not exactly the same group of people of this study, but from the same village) to judge the age range (older vs. younger) of two female speakers in a forced-choice task. The listeners were at chance at estimating the age of the speakers, and the perceived age of the voices had no effect on their cue selection. Therefore, it is unlikely that listeners in this current study were biased by the age of the voices. Table 7 summarizes the F1, F2, and f0 values of the original tokens and the synthesis settings for these three parameters.

There were 180 resynthesized stimuli for each speaker. Fig. 6 illustrates the F1, F2, and f0 manipulations for one token as a cube in a 3-dimensional acoustic space. For each minimal pair, two such cubes were created, one based on the tense register syllable, and the other based on the lax register syllable. Altogether, the stimuli varied in four acoustic dimensions: phonation (tense vs. lax), F1, F2, and f0.

Acoustic measurements of phonation were taken with VoiceSauce in order to ensure that the synthetic stimuli maintained the natural phonation contrast. Paired t-tests were used to test the significance of each measure. The resynthesized tense and lax stimuli remained distinct in several phonation measures. The tense and lax male /be/ and /bu/ stimuli were significantly different in $H1^{*}$-$H2^{*}$, $H1^{*}$-$A1^{*}$, $H1^{*}$-$A2^{*}$, $H1^{*}$-$A3^{*}$, and CPP. The female tense and lax /be/ stimuli were significant different in the measures $H1^{*}$-$H2^{*}$, $H1^{*}$-$A1^{*}$, and $H1^{*}$-$A3^{*}$. The female /bu/ stimuli reached significance in $H1^{*}$-$H2^{*}$, $H1^{*}$-$A2^{*}$, $H1^{*}$-$A3^{*}$, and CPP. Therefore, the manipulation was successful in retaining the phonation distinction between the minimal pairs. Moreover, the intended register categories of the endpoints of the stimuli were accurately identified by the listeners (87% correct for /bu/ syllables, and 93% correct for /be/ syllables), suggesting that the stimuli were good representatives of the tense and lax categories.

2.4.2. Methods

The experimental procedure was the same as Experiment 1. The participants were divided into two groups, 21 participants heard the resynthesized stimuli of the male speaker, and 20 heard the resynthesized stimuli of the female speaker. Each group heard 180 resynthesized stimuli in total, and randomization was carried out over all 180 stimuli. When each stimulus was presented, the participant was asked to choose between two options in Mandarin, one corresponding to the word with tense phonation in the minimal pair, and the other corresponding to word with lax phonation in the minimal pair.

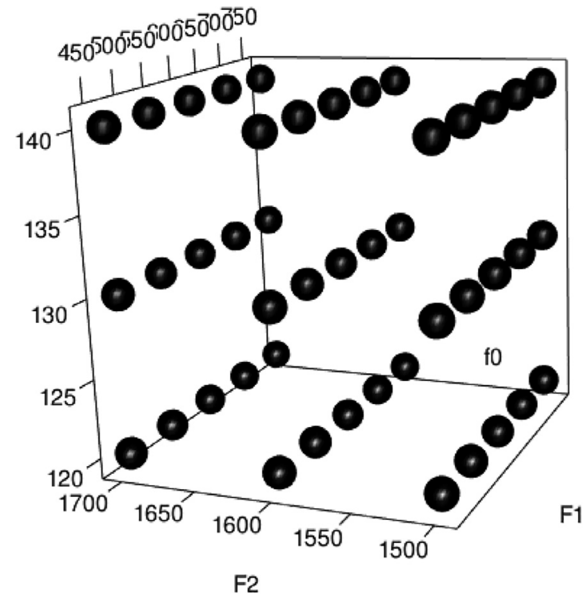


Fig. 6. Demonstration of the manipulation of F1 (five steps), F2 (three steps), and f0 (three steps).

2.4.3. Results

To estimate the relative importance of F1, F2, f0, and phonation in the production of the register contrast, and whether the relative contribution of these acoustic cues was significantly modulated by age groups (old, middle and young) and voice of the stimuli (female voice vs. male voice), a mixed-effects logistic regression model was built for each syllable type (e.g., /be/ or /bu/). The dependent variable was the binary tense vs. lax register contrast, and the fixed factors were F1, F2, f0, phonation, and the interactions of each acoustic cue with age groups and stimuli voice. All dimensions were standardized before entering the models so that the coefficients of different cues are comparable (cf. Schertz et al., 2015). Listeners were included as the random intercept. The models failed to converge when random slopes were added, so only random intercepts were retained. Again, these models were dummy-coded, and the old listener group and female voice were set as the reference levels. The results for /be/ and /bu/ are summarized in Tables 8 and 9 respectively.

Table 7
The original values of the tokens chosen for synthesis and the synthesis settings for each of the parameters.

		Male			Female		
		Tense	Lax	Synthesis Settings	Tense	Lax	Synthesis Settings
be33	F1 (Hz)	655	476	450, 525, 600, 675, 750	838	647	600, 675, 750, 825, 900
	F2 (Hz)	1582	1602	1500, 1600, 1700	1768	1977	1700, 1850, 2000
	f0 (Hz)	137	150	120, 130, 140	190	188	178, 188, 198
bu33	F1 (Hz)	334	317	250, 325, 400, 475, 550	536	479	350, 425, 500, 575, 650
	F2 (Hz)	1198	1342	1150, 1250, 1350	1364	1402	1300, 1400, 1500
	f0 (Hz)	117	120	110, 120, 130	192	184	180, 190, 200

Table 8
Summary of the mixed-effects logistic regression models for the perception of /be/ (dummy-coded; reference level for age_group = old listeners; reference level for voice = female voice. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$).

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.70	0.39	1.78	0.07
F1	1.55	0.37	4.18	0.00***
voice	−0.84	0.58	−1.45	0.15
age_group(middle)	−0.19	0.47	−0.40	0.69
age_group(young)	0.30	0.98	0.31	0.76
phonation	0.78	0.13	6.14	0.00***
F2	−0.20	0.12	−1.65	0.10
f0	0.13	0.15	0.90	0.37
F1:voice	−0.09	0.55	−0.17	0.86
F1:age_group(middle)	0.09	0.45	0.21	0.84
F1:age_group(young)	1.08	1.00	1.09	0.28
voice:age_group(middle)	0.22	0.71	0.32	0.75
voice:age_group(young)	−1.44	1.15	−1.25	0.21
voice:phonation	−0.56	0.18	−3.11	0.00**
age_group(middle):phonation	−0.16	0.15	−1.03	0.30
age_group(young):phonation	0.74	0.45	1.65	0.10
voice:F2	0.16	0.18	0.90	0.37
age_group(middle):F2	0.08	0.15	0.57	0.57
age_group(young):F2	−0.35	0.37	−0.95	0.34
voice:f0	−0.13	0.22	−0.61	0.54
age_group(middle):f0	0.09	0.18	0.49	0.62
age_group(young):f0	0.69	0.43	1.60	0.11
F1:voice:age_group(middle)	0.21	0.67	0.32	0.75
F1:voice:age_group(young)	0.54	1.16	0.47	0.64
voice:age_group(middle):phonation	0.40	0.22	1.83	0.07
voice:age_group(young):phonation	−0.20	0.49	−0.40	0.69
voice:age_group(middle):F2	−0.19	0.22	−0.86	0.39
voice:age_group(young):F2	0.36	0.42	0.86	0.39
voice:age_group(middle):f0	0.19	0.27	0.70	0.48
voice:age_group(young):f0	−0.31	0.49	−0.62	0.54

Table 9
Summary of the mixed-effects logistic regression models for the perception of /bu/ (dummy-coded; reference level for age_group = old listeners; reference level for voice = female voice. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$).

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	−0.08	0.44	−0.19	0.85
F1	0.31	0.15	2.07	0.04*
voice	0.00	0.57	0.00	1.00
age_group(middle)	−0.20	0.53	−0.38	0.70
age_group(young)	0.31	0.63	0.49	0.62
phonation	0.94	0.11	8.47	0.00***
F2	−0.04	0.11	−0.39	0.69
f0	−0.04	0.14	−0.31	0.76
F1:voice	−0.07	0.20	−0.34	0.74
F1:age_group(middle)	0.06	0.18	0.36	0.72
F1:age_group(young)	0.60	0.18	3.28	0.00**
voice:age_group(middle)	−0.44	0.73	−0.60	0.55
voice:age_group(young)	−0.13	0.85	−0.16	0.88
voice:phonation	−0.65	0.14	−4.53	0.00***
age_group(middle):phonation	−0.01	0.13	−0.04	0.97
age_group(young):phonation	0.02	0.17	0.14	0.89
voice:F2	0.00	0.14	0.01	0.99
age_group(middle):F2	0.01	0.13	0.08	0.94
age_group(young):F2	0.13	0.16	0.82	0.41
voice:f0	0.38	0.23	1.64	0.10
age_group(middle):f0	0.10	0.17	0.60	0.55
age_group(young):f0	0.24	0.21	1.16	0.25
F1:voice:age_group(middle)	0.08	0.25	0.34	0.74
F1:voice:age_group(young)	−0.03	0.30	−0.09	0.93
voice:age_group(middle):phonation	−0.06	0.18	−0.33	0.74
voice:age_group(young):phonation	0.02	0.22	0.08	0.93
voice:age_group(middle):F2	−0.12	0.18	−0.68	0.50
voice:age_group(young):F2	−0.23	0.21	−1.08	0.28
voice:age_group(middle):f0	0.19	0.23	0.83	0.40
voice:age_group(young):f0	0.49	0.28	1.76	0.08

2.4.3.1. Stimuli voice effects. As shown in Tables 8 and 9, a significant interaction between stimuli voice and phonation was found for both /be/ and /bu/, indicating a perceptual bias between the two voices. This bias can be understood with Fig. 7. For both /be/ and /bu/, the classification of tense vs. lax registers is slightly better (i.e., larger difference in the percentage of "lax" responses between tense and lax stimuli) with the female voice, probably because the phonation distinction is slightly greater for the female speaker than for the male speaker (Euclidean distances between the tense and lax phonation measures were calculated; /be/: male 2.47 vs. female 3.07, and /bu/: male 2.42 vs. female 2.52). Since we did not explicitly control for phonation cues, this type of bias could not be avoided. Crucially, there are no interactions between voice with any other cues, such as F1, and there are also no interactions between age group and voice. This means that although voice can slightly bias the listeners' judgment of the range of tenseness, different voices did not affect listeners' cue weighting. Therefore, we combined the voice groups for the rest of the analysis.

2.4.3.2. Age effects. For syllable /be/, there are no significant interactions between age group and any of the phonetic cues, indicating that all age groups had the same cue weighting for the classification of the register contrast. Additionally, while both F1 and phonation significantly contribute to the perceptual classification, F1 carries a higher weight (cf. Fig. 14). There are no significant effects or interactions with age for F2 and f0. These results can be understood in Figs. 8–10, which illustrate the percentage of "lax" responses as the result of different cue steps for each acoustic dimension. Overall, the listeners generally had more "lax" responses for lax stimuli, indicating that they were sensitive to phonation cues. There is also a contrast in the importance of F1 and the other coarticulated cues: As shown in Fig. 8, the percentage of "lax" responses decreases substantially as F1 increases, but F2 (Fig. 9) and f0 (Fig. 10) steps have little effects on the percentage of "lax" responses. Overall, there are no age effects on cue weighting for /be/.

For syllable /bu/, as indicated by the coefficients in Table 9, phonation is the primary cue for classifying the registers, and there is no interaction between phonation and age group. In addition to phonation, F1 is also a significant predictor for the registers. Importantly, the interaction between F1 and young age group is significant, indicating that the young speakers put more weight on F1 than the old speakers. Again, there are no significant effects or interactions with age for F2 and f0. This pattern can be understood in Figs. 11–13, which again depict the percentage of "lax" responses as the result of different cue steps along each acoustic dimension. Overall, listeners mostly relied on the phonation of the stimuli to classify the tense vs. lax registers. As shown in Fig. 11, unlike /be/ (cf. Fig. 8), F1 plays a much weaker role in the register classification for /bu/. However, the effect of F1 is noticeably stronger for the young speakers, as the change of "lax" responses per step is greater and more consistent for the young speakers.

Taken altogether, this experiment shows that while listeners mostly relied on formants for the register classification of the syllable /be/, they relied more on phonation for /bu/. Listener age group differences can also be observed. For /be/, all age groups consistently relied on F1 to perceive the register contrast, and phonation cues only played a secondary role in perception. For /bu/, however, the classification of the register

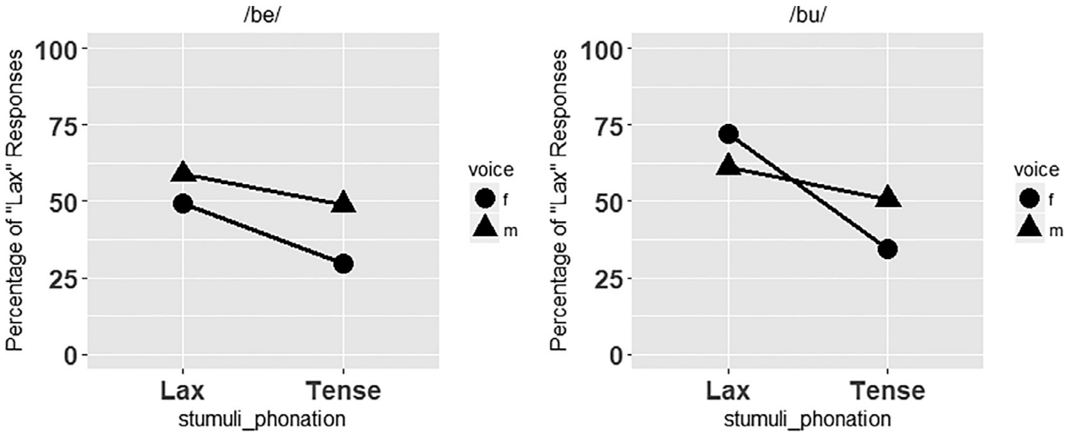


Fig. 7. Percentage of “lax” responses based on different voices.

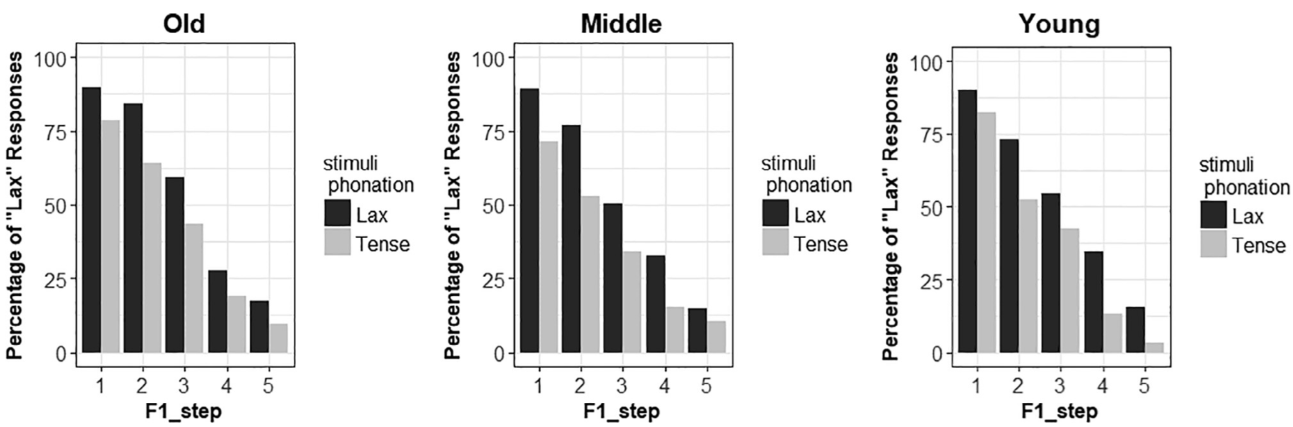


Fig. 8. Syllable /be/: Percentage of “lax” responses along the F1 continuum.

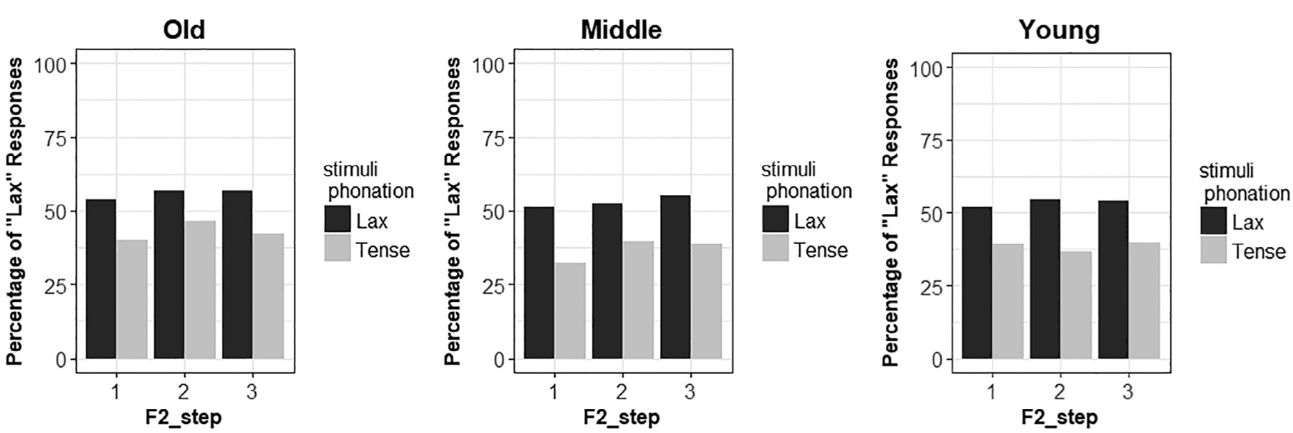


Fig. 9. Syllable /be/: Percentage of “lax” responses along the F2 continuum.

contrast depended more on phonation, though the three age groups did not behave in the same way. The oldest group still overwhelmingly relied on phonation cues to perceive the register contrast, but the youngest group weighed F1 more heavily.

2.5. Mapping between production and perception

2.5.1. Mapping at the group level

To understand how the association between register, acoustic parameters, and age differ between speech production and perception, and to more straightforwardly compare the relative cue weighting between production and perception within each age group, a series of mixed-effects logistic regression models were run for each syllable type and each age group. For all the models, the binary dependent variable was the tense vs. lax registers, and the four competing phonetic cues (phonation, F1, F2, and f0) were the fixed factors, and subjects were included as the random effects. All variables were standardized before being included in the models. In each regression model, the relative importance of each cue was estimated

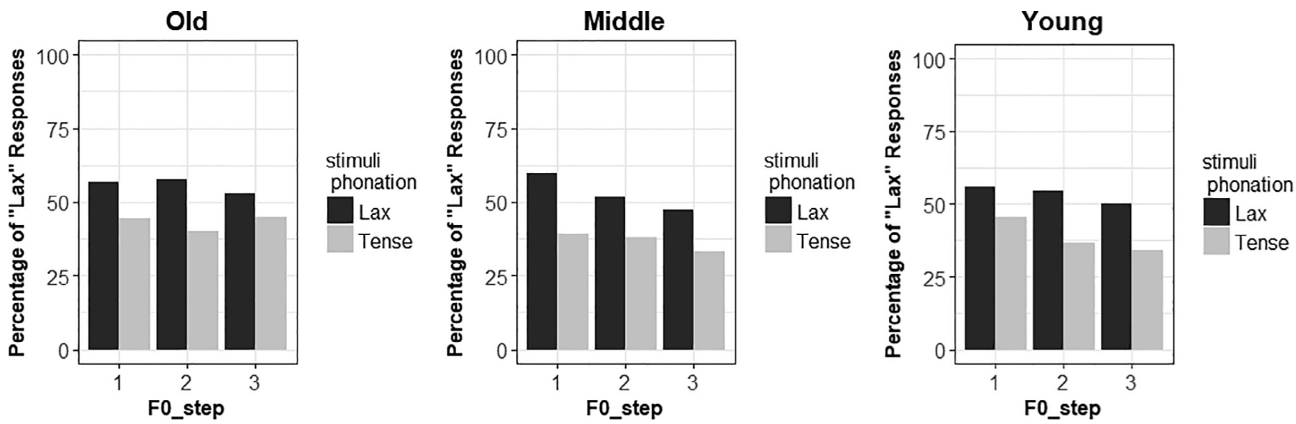


Fig. 10. Syllable /be/: Percentage of “lax” responses along the f0 continuum.

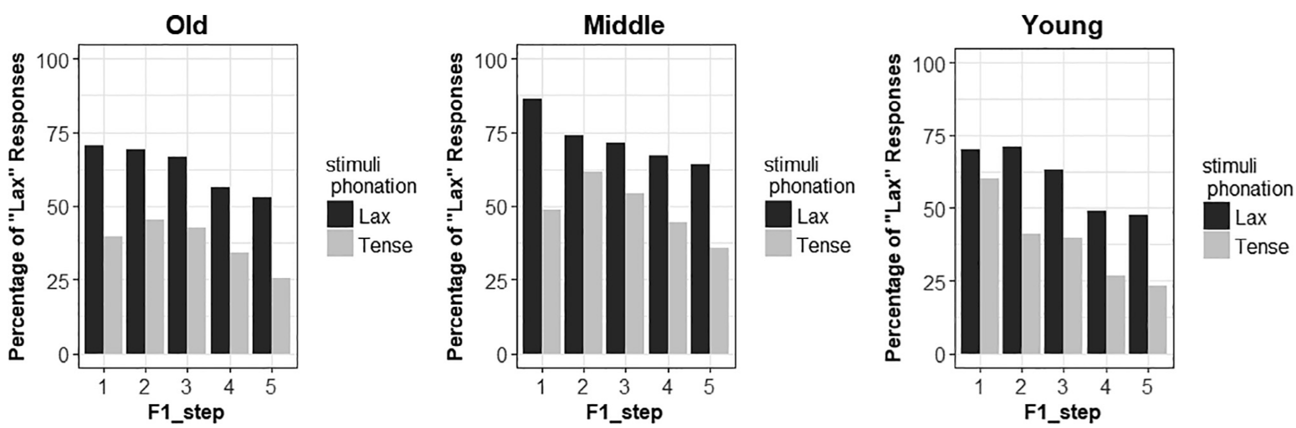


Fig. 11. Syllable /bu/: Percentage of “lax” responses along the F1 continuum.

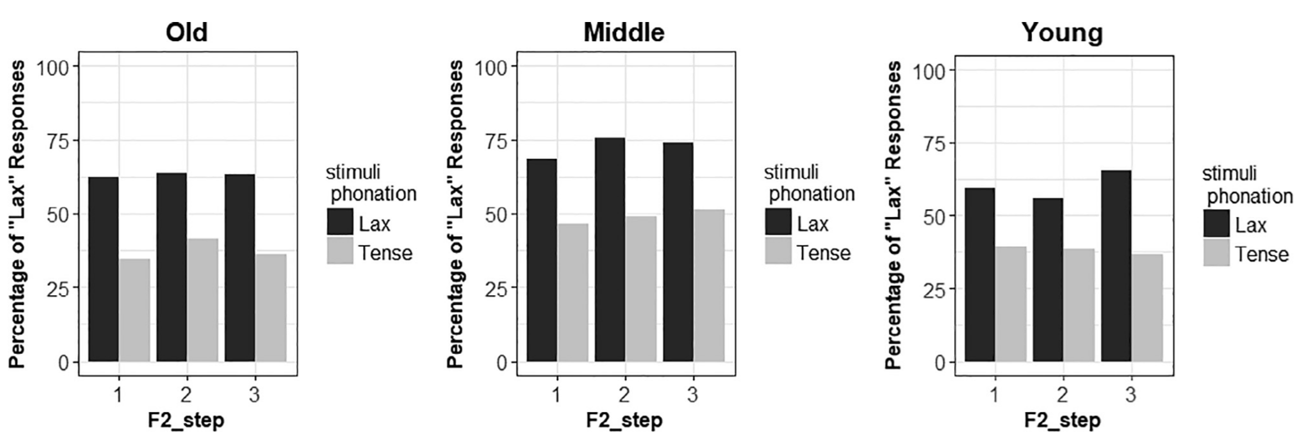


Fig. 12. Syllable /bu/: Percentage of “lax” responses along the F2 continuum.

by the coefficients. Greater absolute values of the coefficients indicate greater importance. We visualize the relative importance of the cues in Fig. 14 (/be/) and Fig. 15 (/bu/), and the significant cues are marked with stars. Detailed output can be found in Appendix 1.

The upper panel of Fig. 14 can be cross-validated with the production pattern discussed in 2.3.2. For /be/, there is a clear difference in the relative cue weighting among different age groups for the register contrast. For the older age groups, while phonation is the most important acoustic cue, F1 (and also F2) also contribute heavily to the register contrast. On the other hand, the youngest group has stopped using phonation as a cue in this contrast and relies solely on F1 and F2. This suggests that the primary cues produced by native speakers of Southern Yi are changing in apparent time for /be/. The lower panel of Fig. 14 can be cross-validated with the perception pattern discussed in 2.4.3. All age groups consistently relied on F1 as the primary cue to perceptually classify the register contrast.

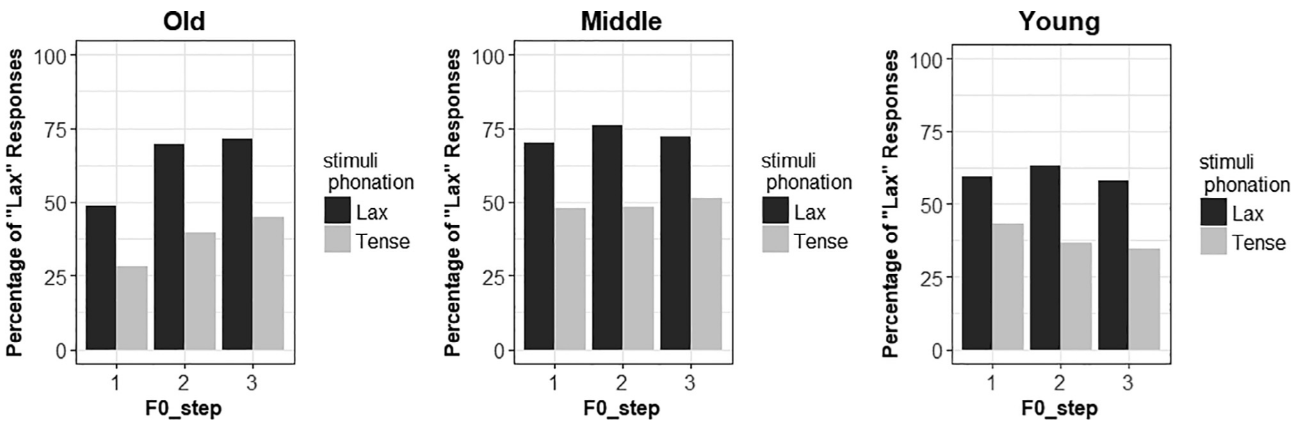


Fig. 13. Syllable /bu/: Percentage of “lax” responses along the f0 continuum.

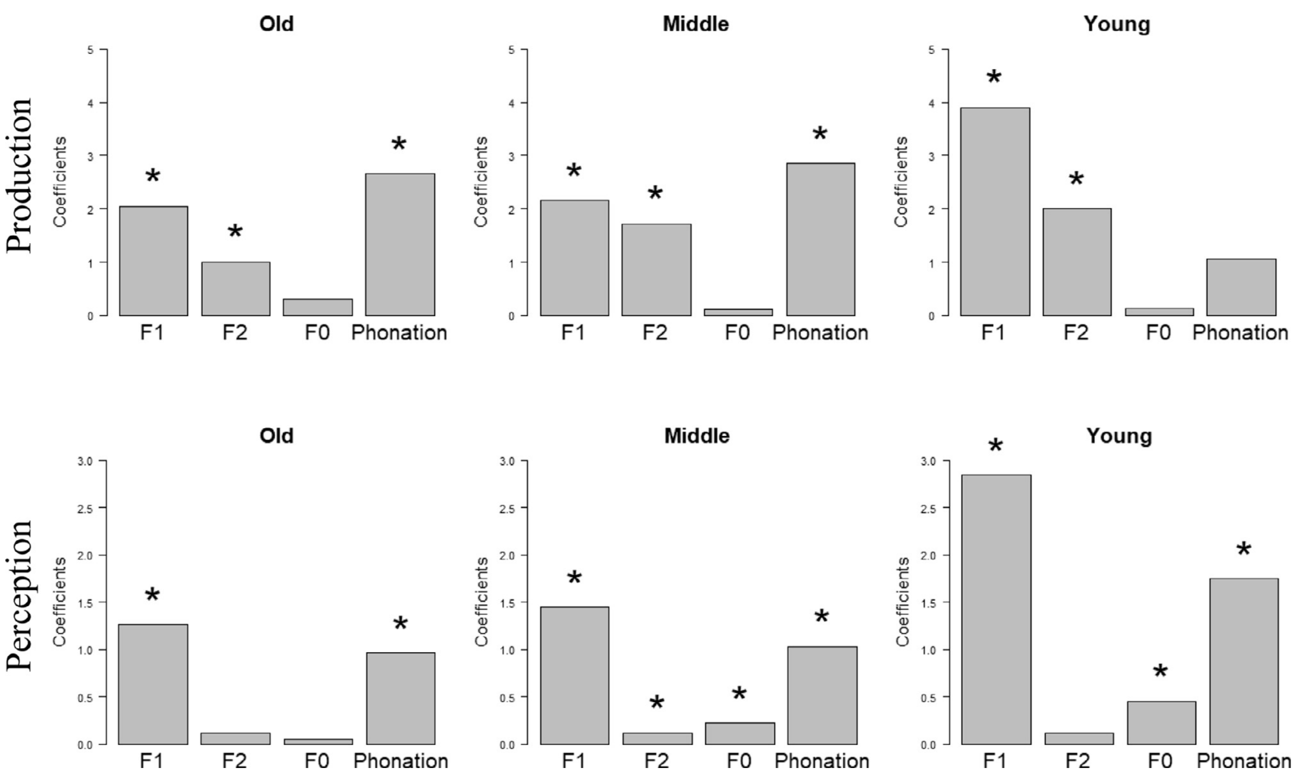


Fig. 14. Relative cue weights for /be/. Production is the upper panel; perception is in the lower panel. Age groups are organized from left to right: old, middle and young. Significant cues are marked with asterisks.

Fig. 15 depicts the relative cue weighting in production and perception for the syllable /bu/. Again, consistent with the results from 2.3.2, all three age groups consistently used phonation as the primary cue in production. However, there are age differences in the cue weighting for perception. As can be seen in Fig. 15 lower panel (also cf. Section 2.4.2), while phonation remains the primary cue of the perception, the relative importance of F1 is growing over time; for the youngest group, the weight of F1 has increased relative to phonation.

A striking pattern can be observed here: the primary cues in production and perception are not always aligned among different groups of speakers. For /be/, in production, the primary cue is in the process of shifting from phonation to formants; however, in perception, the shift is already complete, and F1 is the primary cue for all age groups. As a result, for the oldest speakers, production and perception are misaligned since they use different primary cues in perception and production, with perception being more advanced in this sound change in progress. By contrast, for /bu/, all age groups are consistent in their cue weighting in production, with phonation as the primary cue, and F1 is insignificant. In perception, however, the relative importance of F1 is increasing over apparent time.

2.5.2. Mapping at the individual level

While we have discussed mapping at the group level, the question remains whether the alignment and misalignment between production and perception can also occur at the

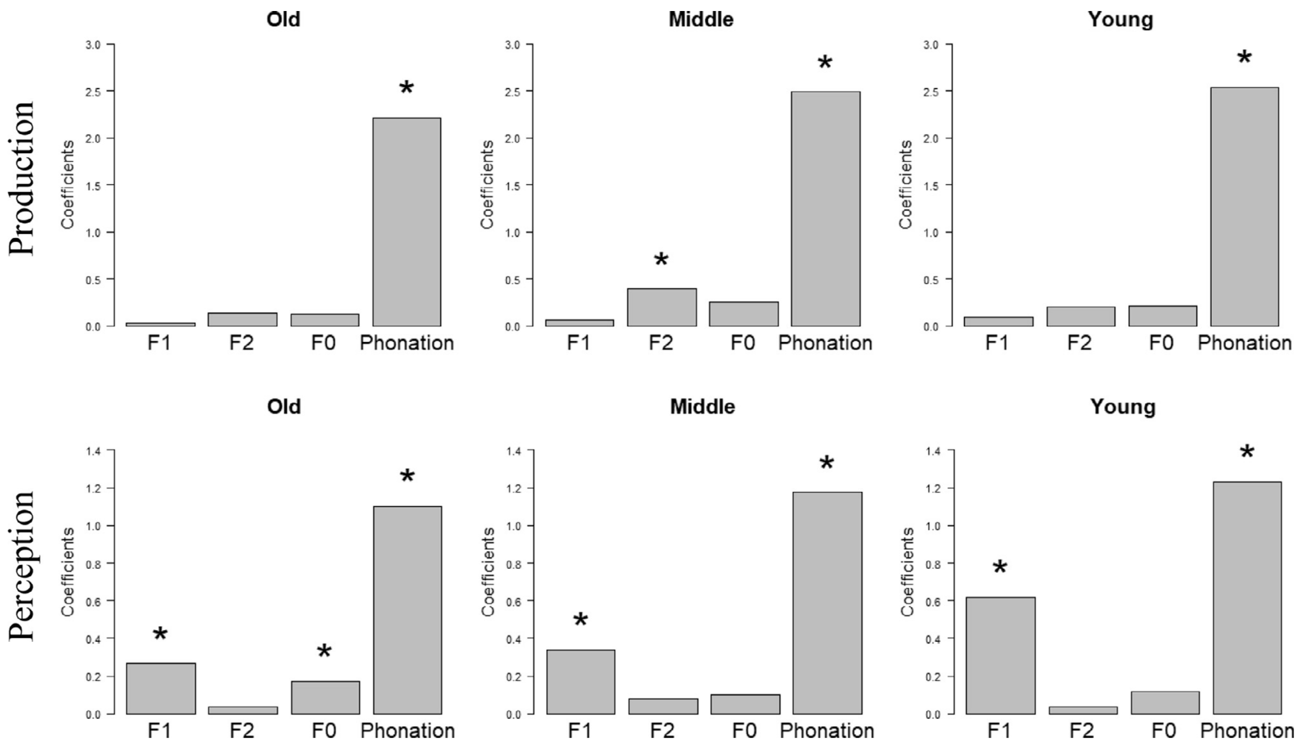


Fig. 15. Relative cue weights for /bu/. Production is the upper panel; perception is in the lower panel. Age groups are organized from left to right: old, middle, and young. Significant cues are marked with asterisks.

individual level. A tight link between production and perception at the individual level is usually assumed (e.g., Beddor, 2009; Lindblom, Guion, Hura, Moon, & Willerman, 1995; Ohala, 1981), but it is hard to find perfect correlations between the production and perception performances from the same speaker to support such a link (e.g., Grosvald & Corina, 2012; Schertz et al., 2015; Shultz, Francis, & Llanos, 2012). One important reason for this is that perception is highly flexible and adaptive, causing the production-perception link to be complex and multifaceted (e.g., Beddor, Harnsberger, & Lindemann, 2002; Sonderegger & Yu, 2010; Zellou, 2017). Another possible reason is that the link between production and perception is rather abstract, so the linear correlations between production and perception of individual cues (e.g., Schertz et al., 2015; Shultz et al., 2012) may be not the best measure. Since the relative importance of a cue is affected by the contributions of other cues, when different speakers produce this cue the same way, it does not follow that this cue has the same relative importance for these speakers. For example, in our case, the increasing importance of F1 for the young speakers in the production cue-weighting model does not necessarily mean that these speakers produce larger F1 differences (cf. Section 2.3.2). Therefore, here we use a relatively integrative and abstract approach to examine the alignment between production and perception at the individual level.

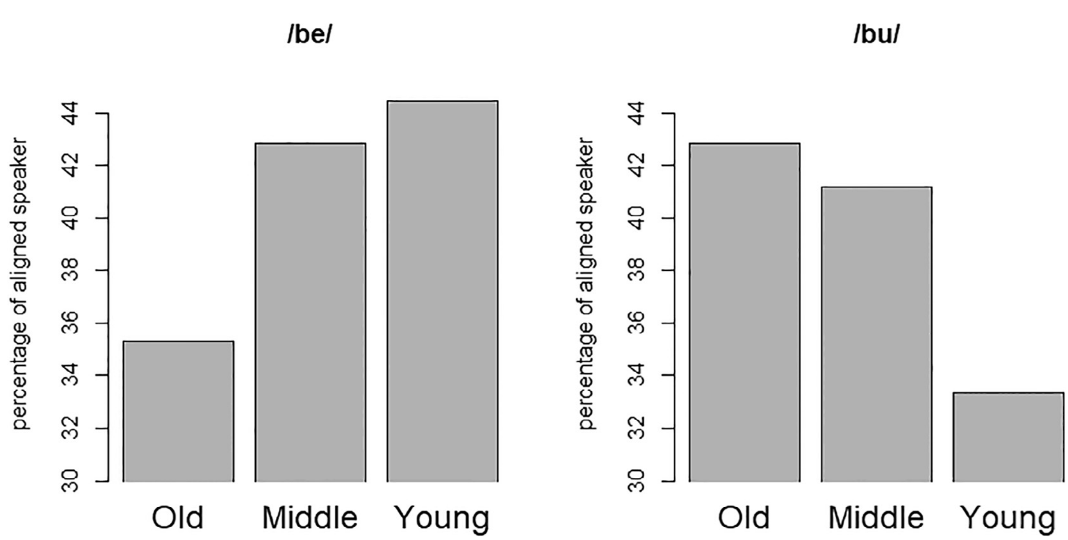


Fig. 16. Percentages of aligned speakers among three age group.

Logistic regression models for both production and perception were first run for each speaker individually. Next, the coefficients of the four relevant cues (F1, F2, f0, and phonation) from both production and perception models were extracted and ranked, and /be/ and /bu/ were calculated separately. If a speaker has identical primary cues for both production and perception, he/she is denoted as an "aligned speaker"; if a speaker has different primary cues for production and perception, he/she is denoted as a "misaligned speaker". We then calculated the percentages of "aligned" and "misaligned" speakers in each age group. As shown in Fig. 16, while none of the age groups is made up of only aligned or misaligned speakers, an interesting contrast between /be/ and /bu/ can be observed. For /be/, the percentage of aligned speakers is higher for the young speakers but much lower for the old speakers; for /bu/, the percentage of aligned speakers is much lower for the young speakers but much higher for the old speakers. Therefore, for /be/, speakers are moving towards an alignment between production and perception; on the other hand, for /bu/, speakers are moving away from an alignment between production and perception.

3. Discussion

3.1. A change in progress

In this study, production and perception experiments were conducted with a single group of Southern Yi speakers to examine the relative cue weighting of multiple co-varying cues in the language's register contrast. In both perception and production, the speakers were able to maintain the tense and lax registers as separate categories. However, as shown in the production experiment, while older speakers used phonation as the primary cue of the register contrast, the youngest group used F1 as the primary cue for the register contrast of the /be/ syllables. For the /bu/ syllables, on the other hand, there was no difference in the production among the three age groups. The finding of the age difference in production generally confirmed the first author's previous observation that sound change is underway – while the register contrast is maintained, formants are overtaking phonation as the primary cues for non-high vowels.

However, the age difference in production does not provide the full picture of the sound change because it alone does not provide sufficient information about when and how the sound change was initiated. In order to gain a better understanding of the time course of this sound change, we need to take into account both production and perception differences among the different age groups. When the cue weights in production and perception of the same groups of people are compared, strikingly, production and perception are misaligned for some speaker groups. For /be/, the perception and production of the older groups are misaligned (Fig. 14), and for /bu/, the perception and production of all groups are misaligned (Fig. 15). This misalignment indicates that cue shifting does not take place in production and perception at the same time. In both cases, the shift occurs in the perception domain earlier than production. Therefore, these results suggest a more nuanced progression of the sound change: For /be/, although the older speakers still primarily rely on phonation cues in production, in perception they have shifted their attention to F1 already; the youngest speakers have shifted to F1 in both production and perception, indicating that they are more advanced in the sound change. What is especially interesting is that even for the high vowel /bu/, although cue weighting remains the same in production, the shift has begun in perception, as the importance of F1 has significantly increased for youngest speakers. Overall, our results suggest a possible time course of cue shifting: perceptual innovation is ahead of production – formant cues have become more dominant in perception before a similar shift occurs in production.

3.2. The mapping between production and perception at different stages of sound change

In Section 1.1, we outlined three possibilities for the time course of sound change: (1) the primary cue shifts in production and perception at the same time, (2) cue shifting occurs in production first, and (3) cue shifting occurs in perception first. Our findings generally support possibility (2): Listeners first shift their attention to the new primary cue, and only later tune their production accordingly. Therefore, when a sound change is initiated, production and perception at the group level tend to be misaligned. The differences in alignment patterns of /be/ and /bu/ suggest that /be/ and /bu/ are at different stages of sound change. The non-high vowel /be/ is at a more advanced stage of sound change. The older speakers are at the initiation stage: The cue weighting shift has occurred in their perception, but they have not tuned their production to the new weighting. For the youngest speakers, the cue shifting is complete in both production and perception. The high vowel in /bu/ is at a much earlier stage of sound change than the vowel in /be/. Although cue shifting has not happened in production at all based on these measures, different age groups appear to have begun to interpret the coarticulatory effects differently. Our results suggest that while older speakers still interpret F1 differences as the coarticulatory effects of distinct phonation, the youngest speakers are entering the initiation stage for /bu/ since they are starting to give up recovering the source of coarticulation.

A similar misalignment in production and perception was reported for Standard Southern British, where /u/-fronting and /ʊ/-fronting are sound changes in progress (Harrington et al., 2008; Harrington, 2012). In both cases, the fronting is caused by coarticulatory effects when the vowel follows a coronal. As the change progresses, the vowels not in this context also undergo a similar fronting. The change for /u/ is more advanced than for /ʊ/ in that younger speakers are shown to shift in both production and perception. However, for /ʊ/, production and perception are misaligned. Perception is ahead of production in the younger age group, and this indicates that /u/-fronting began earlier than /ʊ/-fronting. Harrington (2012) proposed that before and after sound change, the link between production and perception is fairly stable, so an alignment between perception and production can be observed. However, when sound change is in progress, the link between production and perception becomes unstable and misaligned, and innovation gradually takes place first in perception and then in production. Harrington (2012) proposed a model of the stages of sound change, reproduced in Fig. 17. Before and after

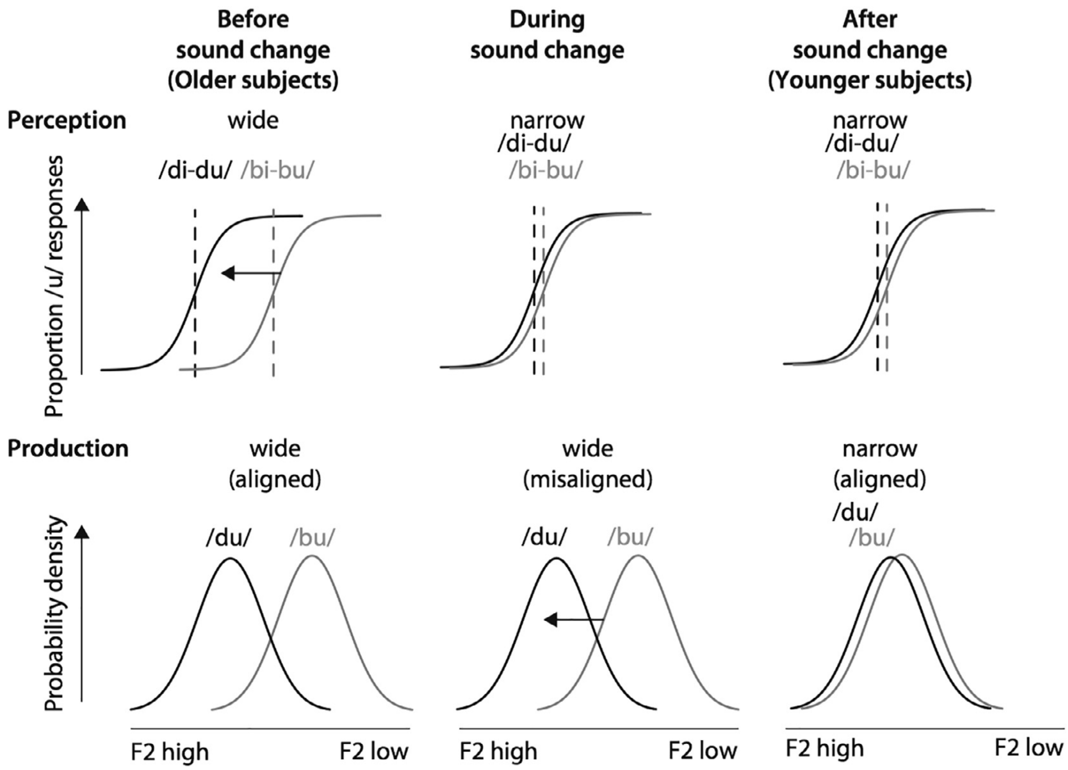


Fig. 17. Stages of sound change (reproduced from Harrington, 2012).

sound change, the phonetic boundary for /di/-/du/ continuum is in alignment between production and perception, but during sound change, production and perception are misaligned, and the shift of the boundary first occurs in perception.

There is a number of differences between our study and Harrington (2012). The vowel split in our study is a different type of sound change than vowel fronting, and we used different experimental paradigms to explore the mapping between production and perception. However, despite these differences, the mapping pattern found in this study is comparable to Harrington (2012). We can generalize the patterns from Figs. 14–18. For /be/, cue shifting appears to be complete in both production and perception for younger speakers, so the youngest speakers are at stage 3. On the other hand, the older speakers still rely on the conservative phonation cue in production, so they are at stage 2. For /bu/, the older speakers can be considered to be at stage 1, as cue shifting has not taken place in production or perception, whereas the youngest speakers appear to approach stage 2, as innovative cues start to enter the perception domain. Thus, in the stable stages, such as stage 1 and stage 3, production and perception are generally aligned (at least at the group level). However, during the initiation stage of a sound change, production and perception become misaligned when the innovative cue first enters the

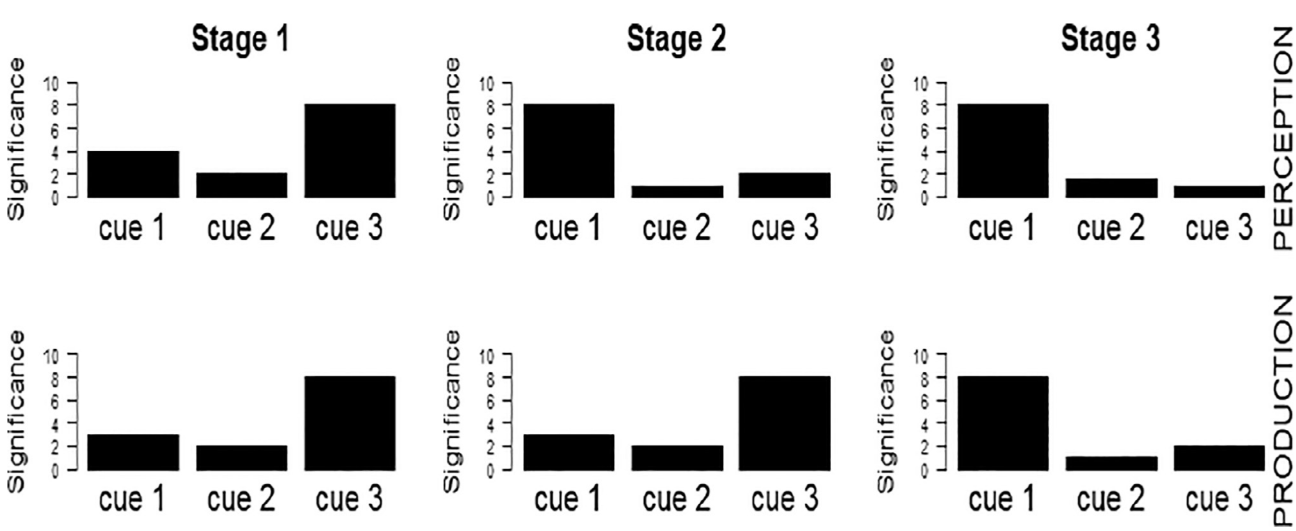


Fig. 18. Hypothetical stages of cue shifting in production and perception.

perception domain before passing into the speaker's production. In Harrington's (2012) model, stage 2 was labeled as "during change", and stage 3 was labeled as "after change". While we agree that the existence of misalignment between production and perception at the group level is a sign of sound change in progress, we did not label the stages as Harrington did, since sound change at the community level covers more than one stage. For example, /be/ is in stage 2 for older speakers and 3 for younger speakers.

The analysis at the individual level provides us with more insights about the production-perception link. Similar to previous studies, the correlations between production and perception are highly variable across speakers, and both aligned and misaligned speakers co-exist in the community. Importantly, it appears that the number of aligned speakers increases when a sound change is close to completion (e.g., /be/), while the number of misaligned speakers increase when a sound change is initiating (e.g., /bu/). Therefore, the alignment between perception and production can be generalized to both the individual level and the group level: Based on our results, cue weights generally match between production and perception in stable stages, like stages 1 and 3; however, once a sound change has initiated, the link between production and perception becomes unstable and loose, and misalignment between production and perception is more likely to happen.

3.3. The role of perception: a more nuanced view

Our findings have important implications for theories of sound change, especially how sound change is initiated through the interaction between production and perception. Our results are generally consistent with the hypothesis that change is initiated when listeners reanalyze coarticulatory effects (i.e., tongue root retraction in our case) in the language input (Beddor, 2009; Ohala 1993, 2005). The vowel split change in Southern Yi occurs probably because younger speakers no longer attribute F1 differences to the phonation distinction, the original source of the register contrast. The cue shifting for syllable /be/ provides an example of how perception may lead the process of reanalysis. As can be seen in Fig. 14, the primary cue for perception has shifted to vowel formants even among the oldest speakers, but the same shift in production is only completed among the youngest speakers. The weakening of phonation cues among the younger speakers is likely due to the process of retuning between production and perception.

The situation for /bu/ is different. In the case of /be/, there is a strong coarticulatory effect between F1 and phonation, so this shift is well grounded in production. As shown in Fig. 4, F1 already sufficiently distinguishes the register contrast even for the older speakers. By contrast, for /bu/, the weight of F1 is rising in perception with each successive age group even though there is no evidence of increased F1 differences in production. In this case, the coarticulatory effect between tense phonation and higher F1 is very weak for /bu/, so this shift is not fully grounded in production. However, it is common for sound changes that originated in particular phonetic contexts to spread and generalize to other contexts. It has been proposed that analogy contributes to the spread of sound change (Kiparsky, 1995). It is possible that the change of perceptual weight of /bu/ is due to analogy, and perception plays an important role in initiating the process of generalization. Therefore, our data suggest that perception can play a leading role in sound change at two stages: (1) the initiation of the reanalysis, where the a coarticulated cue increases in importance relative to the source cue; (2) the generalization of a sound change to other phonetic contexts.

Our finding that cue shifting occurs first in perception differs from the trend reported in Coetzee et al. (2018) and Pinget (2015). However, since perception is flexible and adaptive, the findings of these studies do not necessarily contradict each other. In light of our findings, it is possible for perception and production to have different mappings at different stages of sound change. At the very beginning of a sound change, it is possible for perception to initiate the reanalysis process, as in our case and /u/-fronting in Southern British English (Harrington, 2012; Kleber et al., 2012). When a sound change is close to completion, listeners are likely to retain sensitivity to the old cue for a while, as in the cases of Afrikaans tonogenesis (Coetzee et al., 2018) and Dutch devoicing (Pinget, 2015). In fact, the lag in perception is also observed in our data: As shown in Fig. 14, while the young speakers no longer rely on phonation cues to produce the register contrast for syllable /be/, they are still sensitive to phonation cues in the perception experiment. Therefore, the different directions of the misalignment are likely to be related to the stages of sound change.

3.4. The role of production

It is impressive that the misalignment of production and perception does not create communication difficulty for innovative and conservative listeners/speakers in the same speech community. Of course, in real life communication, there is much more top-down knowledge available, so speakers can rely on contextual cues to recognize the intended words. However, even in a monosyllabic identification task (as in our first perception experiment), where little contextual information was provided, the shift of cue weighting did not affect the overall perceptibility of the contrast. One possible explanation is that the different perception strategies are all compatible with the input and are thus equally effective. Beddor (2009, 2012) proposed that listeners play an active role in sound change: Listeners can choose between alternative perception strategies since different weightings of the coarticulatory source and its effects can be compatible with the same input. Innovative listeners would choose to put more weight on the coarticulatory effects while conservative listeners would weigh the source of coarticulation more heavily. For example, in American English VNC (e.g. spend, spent) words, the duration of the nasalized vowel and the following nasal consonant have a trade-off relationship: A vowel with a longer nasalized portion tends to be followed by a short nasal consonant, as in *spent*, and a vowel with a briefly nasalized portion tends to be followed by a long nasal consonant, as in *spend*. Because the overall amount of nasalization usually remains constant, listeners are able to treat these two cues as equivalent sources of nasality. The perceptual equivalence of nasality on either V or N also preserves coda nasals from loss, since N can still be perceived even with relatively little nasality. A similar explanation can be adopted here: Listeners are actively selecting which cues to attend to, and cue shifting is especially likely to happen when a co-varying relationship between the source and its effects is established. Because there is a strong co-variation between phonation and F1, attending to either or both cues results in the same category classification. To validate this proposal, a further correlation analysis was carried out for /be/ and /bu/ between the phonation cues and F1 in production. H1*-H2* has a significant correlation, or co-varying relationship, with F1 for /be/ ($r = 0.46$, $p < 0.01$), but there is no significant correlation between H1*-H2* and F1 for /bu/ ($r = 0.08$, $p > 0.05$) (c.f. Appendix 2). Therefore, cue shifting between phonation cues and formants is more likely for /be/ than for /bu/ given the input. Indeed, the sound change takes place first for the non-high vowel. Due to the significant co-varying relationship between the phonation cues and vowel quality cues for /be/, listeners have the opportunity to analyze F1 as equally informative as phonation. Therefore, innovative listeners are able to actively switch attention to the vowel quality cues without compromising the perceptibility of the contrast. Furthermore, these correlations also explain why other coarticulatory cues, such as f0 (cf. Kuang, 2013), do not participate in cue shifting the same way as F1, because these cues do not have co-varying relationship with phonation. Since co-variation is a facilitating condition for cue shifting to happen, further studies are needed to better understand how a co-varying relationship between cues is established.

4. Conclusion

Based on examining the relative cue weighting in production and perception of the register contrast in Southern Yi, our results suggest that the register contrast in Southern Yi is undergoing sound change: While speakers of all age groups still maintain this contrast, the primary cue of this contrast is shifting, and vowel quality is overtaking phonation as the primary cue. Additionally, the results from the cue weighting experiments indicate that this sound change is more advanced in non-high vowels than high vowels in both perception and production. More importantly, for both non-high vowels and high vowels, production and perception are misaligned. In both cases, the shift from phonation to formant values occurs first in perception, and production lags behind. We proposed that production and perception have more nuanced roles in the cue shifting type of sound change. These findings overall provide a better understanding of the time course of sound change, especially at the initiation stage.

Acknowledgements

This study is supported by a URF award of University of Pennsylvania to Jianjing Kuang. We are grateful to Yan Lu for her help with the fieldwork, and we would like to thank all the Yi friends who kindly participated in this study. We sincerely thank Taehong Cho, Jonathan Harrington, and two anonymous reviewers for their insightful comments on the earlier version of this paper.

Appendix 1. Detailed results from the regression models ($^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$)

/be/ production: young

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	−3.87	0.78	−4.96	0.00***
F1	3.90	1.05	3.72	0.00***
F2	−2.00	0.88	−2.27	0.02*
f0	−0.14	0.74	−0.19	0.85
Phonation	−1.05	1.03	−1.02	0.31

/be/ production: middle

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	−4.03	1.03	−3.91	0.00***
F1	2.15	0.65	3.31	0.00***
F2	−1.72	0.59	−2.91	0.00**
f0	0.11	0.75	0.15	0.88
Phonation	−2.85	0.70	−4.07	0.00***

/be/ production: old

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	−2.84	0.45	−6.26	0.00**
F1	2.05	0.59	3.45	0.00**
F2	−1.01	0.48	−2.11	0.04*
f0	−0.30	0.30	−0.98	0.33
Phonation	−2.66	0.52	−5.10	0.00***

/bu/ production: young

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.61	0.46	3.46	0.00***
F1	−0.10	0.29	−0.35	0.73
F2	0.20	0.22	0.92	0.36
f0	0.21	0.19	1.10	0.27
Phonation	−2.54	0.51	−4.95	0.00***

/bu/ production: middle

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.38	0.36	3.90	0.00***
F1	−0.06	0.20	−0.31	0.76
F2	0.40	0.19	2.06	0.04*
f0	−0.26	0.19	−1.33	0.18
Phonation	−2.49	0.43	−5.79	0.00***

/bu/ production: old

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.24	0.28	4.41	0.00***
F1	0.05	0.16	0.30	0.77
F2	−0.06	0.19	−0.31	0.75
f0	−0.20	0.14	−1.38	0.17
Phonation	−2.41	0.35	−6.94	0.00***

/be/ perception: young

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	−1.71	0.61	−2.79	0.01^{*}
F1	2.85	0.25	11.42	0.00^{***}
F2	−0.12	0.14	−0.86	0.39
f0	0.46	0.14	3.15	0.00^{*}
Phonation	1.75	0.31	5.63	0.00^{*}

/bu/ perception: young

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	−0.35	0.55	−0.64	0.52
F1	0.62	0.08	7.39	0.00^{***}
F2	−0.04	0.08	−0.49	0.62
f0	0.12	0.08	1.47	0.14
Phonation	1.23	0.16	7.48	0.00^{***}

/be/ perception: middle

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	−0.25	0.19	−1.36	0.17
F1	1.45	0.07	20.05	0.00^{***}
F2	−0.12	0.06	−2.03	0.04^{*}
f0	0.22	0.06	3.74	0.00^{**}
Phonation	1.03	0.12	8.46	0.00^{***}

/bu/ perception: middle

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	−1.03	0.14	−7.29	0.00
F1	0.34	0.05	6.47	0.00^{***}
F2	−0.08	0.05	−1.58	0.11
f0	−0.10	0.05	−1.96	0.05
Phonation	1.18	0.10	11.22	0.00^{***}

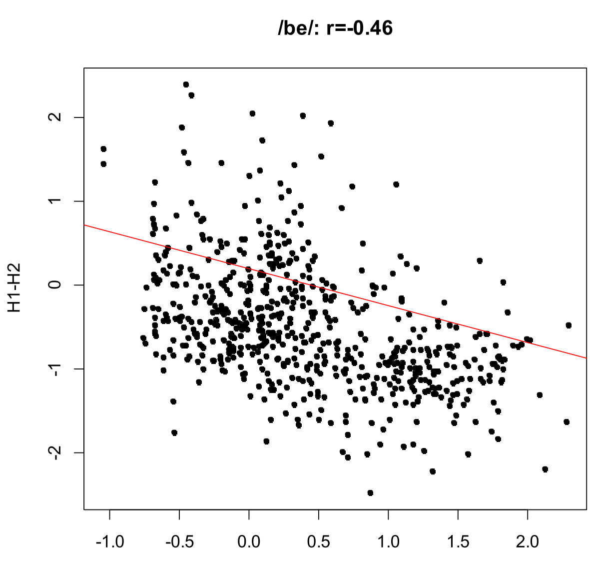
/be/ perception: old

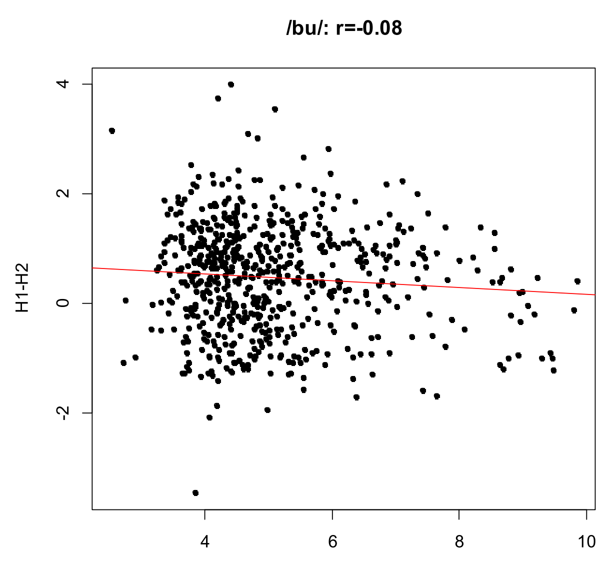
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	−0.27	0.22	−1.21	0.23
F1	1.26	0.10	12.79	0.00^{***}
F2	−0.11	0.08	−1.35	0.18
f0	0.05	0.08	0.62	0.53
Phonation	0.96	0.17	5.55	0.00^{***}

/bu/ perception: old

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	−0.62	0.26	−2.37	0.02^{*}
F1	0.27	0.07	3.88	0.00^{***}
F2	−0.04	0.07	−0.58	0.56
f0	−0.17	0.07	−2.52	0.00^{**}
Phonation	1.10	0.14	8.00	0.00^{***}

Appendix 2. Correlation between H1*-H2* and F1 for /be/ and /bu/





References

Ohala, J. J. (2005). Phonetic explanations for sound patterns. A figure of speech: A Festschrift for John Laver (pp. 23–38).

Sonderegger, M., & Yu, A. (2010). A rational account of perceptual compensation for coarticulation. In Proceedings of the 32nd annual meeting of the cognitive science society.

Abramson, A. S. (2004). The plausibility of phonetic explanations of tonogenesis. From traditional phonology to modern speech processing: Festschrift for Professor Wu Zongji's 95th Birthday (pp. 17–29).

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Bang, H. Y., Sonderegger, M., Kang, Y., Clayards, M., & Yoon, T. J. (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.

Bates, D., Maechler, M., & Bolker, B. (2013). lme4: Linear mixed-effects models using S4 classes, R package version 0.999999-2. http://cran.r-project.org/package=lme4.

Beddor, P. S. (2009). A coarticulatory path to sound change. *Language, 85*(4), 785–821.

Beddor, P. S. (2012). Perception grammars and sound change. In M.-J. Solé & D. Recasens (Eds.), *The initiation of sound change. Perception, production, and social factors* (pp. 37–55). Amsterdam: John Benjamin.

Beddor, P. S., Harnsberger, J. D., & Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics, 30*(4), 591–627.

Brunelle, M. (2012). Dialect experience and perceptual integrity in phonological registers: Fundamental frequency, voice quality and the first formant in Cham. *The Journal of the Acoustical Society of America, 131*(4), 3088–3102.

Chodroff, E., & Wilson, C. (2014). Burst spectrum as a cue for the stop voicing contrast in American English. *The Journal of the Acoustical Society of America, 136*(5), 2762–2772.

Coetzee, A. W., Beddor, P. S., Shedden, K., Styler, W., & Wissing, D. (2018). Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics, 66*, 185–216.

Davidson, L. (2016). Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics, 54*, 35–50.

Edmondson, J. A., & Esling, J. H. (2006). The valves of the throat and their functioning in tone, vocal register and stress: Laryngoscopic case studies. *Phonology, 23*(2), 157–191.

Edmondson, J., Esling, J., Harris, J. G., Shaoni, L., & Ziwo, L. (2001). The aryepiglottic folds and voice quality in the Yi and Bai languages: Laryngoscopic case studies. *Mon-Khmer Studies*, 83–100.

Fowler, C. A. (2005). Parsing coarticulated speech in perception: Effects of coarticulation resistance. *Journal of Phonetics, 33*, 199–213.

Fowler, C. A., & Smith, M. (1986). Speech perception as "vector analysis": An approach to the problems of segmentation and invariance. In J. S. Perkell & D. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 126–136). Hillsdale, NJ: Erlbaum.

Francis, A. L., Kaganovich, N., & Driscoll-Huber, C. (2008). Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English. *The Journal of the Acoustical Society of America, 124*(2), 1234–1251.

Grosvald, M., & Corina, D. (2012). The production and perception of sub-phonemic vowel contrasts and the role of the listener in sound change. In M.-. J. Solé & D. Recasens (Eds.), *The initiation of sound change. Perception, production, and social factors* (pp. 77–100). Amsterdam: John Benjamins Publishing Company.

Haggard, M., Ambler, S., & Callow, M. (1970). Pitch as voicing cue. *The Journal of the Acoustical Society of America, 47*, 613–617.

Haggard, M., Ambler, S., & Callow, M. (1970). Pitch as a voicing cue. *The Journal of the Acoustical Society of America, 47*(2B), 613–617.

Halekoh, U., & Højsgaard, S. (2014). A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models – the R Package pbkrtest. *J. Stat. Soft., 59*(9).

Harrington, J. (2012). The coarticulatory basis of diachronic high back vowel fronting. In M.-. J. Solé & D. Recasens (Eds.), *The initiation of sound change. Perception, production, and social factors* (pp. 103–122). Amsterdam: John Benjamins Publishing Company.

Harrington, J., Kleber, F., & Reubold, U. (2008). Compensation for coarticulation, /u/-fronting, and sound change in Standard Southern British: An acoustic and perceptual study. *Journal of the Acoustical Society of America, 123*, 2825–2835.

Hombert, J. M., Ohala, J. J., & Ewan, W. G. (1979). Phonetic explanations for the development of tones. *Language*, 37–58.

Hyman, L. M. (1976). Phonologization. In A. Juilland (Ed.), *Linguistic studies presented to Joseph H. Greenberg* (pp. 407–418). Saratoga: Anma Libri.

Janson, T. (1983). Sound change in perception and production. *Language*, 18–34.

Jun, S. A. (1996). Influence of microprosody on macroprosody: A case of phrase initial strengthening. *UCLA Working Papers in Phonetics* (pp. 97–116).

Kang, K. H. (2009). *Clear speech production and perception of Korean stops and the sound change in Korean stops* (Doctoral dissertation). Eugene: University of Oregon.

Kang, Y. (2014). Voice onset time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics, 45*, 76–90.

Kang, K. H., & Guion, S. G. (2008). Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. *The Journal of the Acoustical Society of America, 124*(6), 3909–3917.

Keating, P., Esposito, C., Garellek, M., Khan, S. U. D., & Kuang, J. (2011). Phonation contrasts across languages. *Proceedings of the ICPhS XVII*.

Kingston, J. (2005). The phonetics of Athabaskan tonogenesis. In S. Hargus & K. Rice (Eds.), *athabaskan prosody* (pp. 137–184). Amsterdam: John Benjamins.

Kingston, J., Diehl, R. L., Kirk, C. J., & Castleman, W. A. (2008). On the internal perceptual structure of distinctive features: The [voice] contrast. *Journal of Phonetics, 36*, 28–54.

Kiparsky, K. (1995). The phonological basis of sound change. In J. Goldsmith (Ed.), *Handbook of phonological theory* (pp. 640–670). Oxford: Blackwell.

Kirby, J. (2013). The role of probabilistic enhancement in phonologization. In A. Yu (Ed.), *Origins of sound change: Approaches to phonologization* (pp. 228–246). Oxford: Oxford University Press.

Kleber, F., Harrington, J., & Reubold, U. (2012). The relationship between the perception and production of coarticulation during a sound change in progress. *Language and Speech, 55*, 383–405.

Kong, E. J., & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics, 59*, 40–57.

Kuang, J. (2011a). *Production and perception of the phonation contrast in Yi* (Master thesis). Los Angeles: University of California.

Kuang, J. (2013). The tonal space of contrastive five level tones. *Phonetica, 70*(1–2), 1–23.

Kuang, J., & Keating, P. (2014). Glottal articulations in tense vs. lax phonation contrasts. *Journal of the Acoustical Society of America, 136*(5), 2784–2797.

Kuang, J. (2011). Phonation contrast in two register contrast languages and its influence on vowel and tone. Proceedings of the 17th international congress of phonetics sciences (pp. 1146-1149).

Kuang, J., & Cui, A. (2018). Relative cue weighting in production and perception of an ongoing sound change in Southern Yi. *Journal of Phonetics, 71*, 194–214.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests in linear mixed effects models, R package version 2.0-33. https://cran.r-project.org/web/packages/lmerTest/index.html.

Labov, W. (1994). *Principles of linguistic change*. Oxford: Blackwell.

Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1958). Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech, 1*(3), 153–167.

Lindblom, B., Guion, S., Hura, S. L., Moon, S.-J., & Willerman, R. (1995). Is sound change adaptive? *Rivista di Linguistica, 7*, 5–37.

Lisker, L. (1986). "Voicing" in English: A catalogue of acoustic features signaling [b] versus [p] in trochees. *Language and Speech, 29*, 3–11.

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

Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. Proceedings of the 6th ICPhS (pp. 563–567).

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America, 49*(2B), 606–608.

Löfqvist, A., Baer, T., McGarr, N. S., & Seider Story, R. (1989). The cricothyroid muscle in voicing control. *Journal of the Acoustical Society of America, 85*, 1314–1321.

Maddieson, I. (1984). The effects on f0 of a voicing distinction in sonorants and their implications for a theory of tonogenesis. *Journal of Phonetics, 12*(1), 9–15.

Maddieson, I., & Ladefoged, P. (1985). "Tense" and "lax" in four minority languages of China. *UCLA Working papers in Phonetics, 60*, 59–83.

Nelson, Noah Richard, & Wedel, Andrew (2017). The phonetic specificity of competition: Contrastive hyperarticulation of voice onset time in conversational English. *Journal of Phonetics, 64*, 51–70.

Ohala, J. J. (1973). Explanations for the intrinsic pitch of vowels. In *Monthly internal memorandum, phonology laboratory* (pp. 9–26). University of California at Berkeley.

Ohala, J. J. (1981). The listener as a source of sound change. In C. Masek, R. Hendrick, & M. F. Miller (Eds.), *Papers from the parasession on language and behavior* (pp. 178–203). Chicago: Chicago Linguistic Society.

Ohala, J. J. (1993). The phonetics of sound change. Historical linguistics: Problems and perspectives (pp. 237–278).

Pinget, A.-F. (2015). *The actuation of sound change* (Doctoral dissertation). Utrecht: Utrecht University.

R Core Team (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Repp, B. H. (1979). Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech, 22*(2), 173–189.

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics, 52*, 183–204.

Shue, Y.-L., Keating, P., Vicenik, C., & Yu, K. (2011). VoiceSauce: A program for voice analysis. *Proceedings of the ICPhS XVII*, 1846–1849.

Shultz, A. A., Francis, A. L., & Llanos, F. (2012). Differential cue weighting in perception and production of consonant voicing. *The Journal of the Acoustical Society of America, 132*(2), EL95-EL101.

Silva, D. J. (2006). Variation in voice onset time for Korean stops: A case for recent sound change. *Korean Linguistics, 13*(1), 1–16.

Solé, M. J. (2014). The perception of voice-initiating gestures. *Laboratory Phonology, 5*(1), 37–68.

Stevens, K. N., & Keyser, S. J. (2010). Quantal theory, enhancement and overlap. *Journal of Phonetics, 38*(1), 10–19.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 7*(5), 1074.

Thurgood, G. (2002). Vietnamese and tonogenesis: Revising the model and the analysis. *Diachronica, 19*(2), 333–363.

Titze, I. R. (1990). Interpretation of the electroglottographic signal. *Journal of Voice, 4*(1), 1–9.

Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science, 34*(3), 434–464.

Wright, R. (2004). A review of perceptual cues and cue robustness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically-based phonology* (pp. 34–57). Cambridge: Cambridge University Press.

Wright, J. D. (2007). *Laryngeal contrast in Seoul Korean* (Doctoral dissertation). Philadelphia: University of Pennsylvania.

Xu, C. X., & Xu, Y. (2003). Effects of consonant aspiration on Mandarin tones. *Journal of the International Phonetic Association, 33*(2), 165–181.

Zellou, G. (2017). Individual differences in the production of nasal coarticulation and perceptual compensation. *Journal of Phonetics, 61*, 13–29.

Zhang, Z. (2016). Cause-effect relationship between vocal fold physiology and voice production in a three-dimensional phonation model. *The Journal of the Acoustical Society of America, 139*(4), 1493–1507.