Laboratory Phonology
Journal of the Association for Laboratory Phonology

Jiang, B., et al. 2020 Individual and dialect differences in perceiving multiple cues: A tonal register contrast in two Chinese Wu dialects. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 11(1): 11, pp. 1–30. DOI: https://doi.org/10.5334/labphon.266

JOURNAL ARTICLE

Individual and dialect differences in perceiving multiple cues: A tonal register contrast in two Chinese Wu dialects

Bing'er Jiang, Meghan Clayards and Morgan Sonderegger
Linguistics department, McGill University, Montreal, Quebec, CA
Corresponding author: Bing'er Jiang (binger.jiang@mail.mcgill.ca)

This study investigates how multiple cues contribute to multi-dimensional phonological contrasts at both the group level and the individual level, and how dialectal experience shapes listeners' perceptual strategies. We examine a tonal register contrast in two Chinese Wu dialects signaled by three cues: pitch height, voice quality, and pitch contour. We found that 1) at the group level, cue weights are context-specific, i.e., vary by tone, and some contrasts rely more heavily on multiple cues than others; 2) dialectal experience affects listeners' perceptual strategy: Shanghai listeners, with their own dialect having a smaller voice quality distinction, do not rely more on the cue even when listening to stimuli with a clear breathy-modal distinction, comparing to Jiashan listeners; 3) individuals' cue weights are correlated in a positive manner, meaning that some listeners show overall larger cue weights than others; larger variability is found when the contrast has more than one salient cue, in which case individuals have different options of choosing one cue over another as the primary cue and this can work against the positive correlation.

Keywords: Speech perception; cue weighting; individual variability; voice quality; Chinese Wu dialects

1. Introduction

Phonological contrasts are usually signaled by multiple acoustic correlates (see Raphael, 2005 for an overview). In perceiving each contrast, listeners rely on each of these cues to a different extent (e.g., Mayo & Turk, 2004; Francis, Kaganovich, & Driscoll-Huber, 2008; Clayards, 2018). Such inequivalence in the contribution to contrasts is called cue weighting (Holt & Lotto, 2006). A widely studied example is the stop voicing contrast in English, where voiced stops show shorter Voice Onset Time (VOT) and a lower onset F0 while voiceless stops show longer VOT and higher onset F0. Native English speakers primarily use VOT to distinguish voiced stops from voiceless in pre-stress syllable-initial position, with F0 playing a smaller role in perceiving the contrast (Abramson & Lisker, 1985; Gordon, Eberhardt, & Rueckl, 1993; Lisker, 1978; Whalen, Abramson, Lisker, & Mody 1993).

Studies show that cue weights vary as a function of the phonological contrast being signaled and the relative importance of cues also varies with linguistic contexts within a language (Oden & Massaro, 1978). For example, Mayo and Turk (2004) examined the role of VOT and formant frequencies for stop voicing contrasts in different vowel contexts in English. They found that while VOT was always the most important cue, listeners used formant transitions more to distinguish between /ta/ and /da/ than between /ti/ and /di/. For the tonal register contrast in Shanghai Wu, Zhang and Yan (2015) found that F0 onset and voice quality (i.e., breathiness) have different relative weights across different syllable onset manners and different utterance positions.

In the current study, we approach how multiple cues signal multi-dimensional contrasts by examining a tonal register contrast in two Chinese Wu dialects—Jiashan Wu and Shanghai Wu, with a focus on the role of secondary/non-primary cues.[1] In both dialects, three cues are used in signaling the contrast: pitch height (F0 onset), voice quality, and pitch contour (pitch slope). Pitch height is considered the primary cue and the role of voice quality is thought to vary across dialects, playing a weaker role in Shanghai Wu in younger generations (e.g., Gao, 2016). No previous studies have examined perception of Jiashan Wu or quantitatively compared the roles of the different cues to this contrast at the same level of detail in either dialect. Furthermore, while previous studies have examined the role of segmental context in Shanghai Wu (Zhang & Yan, 2015), we compare different tones (contexts) across two experiments. Individual differences are examined in both experiments. By examining these sources of variability (dialect, tone context, and individual) in this multidimensional tonal register contrast, this study aims to answer the following questions: 1) Averaging over listeners within each dialect, how important is the secondary cue (i.e., voice quality) in a multidimensional contrast and does it vary by tone contrast, or is it consistent for different tone pairs? 2) As sound change is taking place in Shanghai, does the reduction of saliency of the voice quality cue result in different cue weighting than for traditional Wu (i.e., Jiashan) listeners? 3) Do individuals show structured differences in cue ordering or cue magnitude, or are differences between individuals random variation? Does the status of the cues in the dialect affect the structure of individual variability?

In subsequent sections, we first discuss the role of secondary cues in individual variability and sound change (1.1); in Section 1.2, we discuss the role of voice quality cross-linguistically. We then give a brief introduction to the two dialects in the study (1.3) and conclude with a more detailed outline of the current study and its methodological contributions (1.4).

1.1. Multiple cues: Individual variability and sound change

Individual variability in speech perception has been well documented (see Yu & Zellou, 2019 for a review). There is also evidence that individuals may vary in their use of secondary cues in perception (see Schertz & Claire, 2019 for a review of individual variability in cue weights). Some individuals use a secondary cue more than others for F0 in English stop voicing (Kapnoula, Winn, Kong, Edwards, & McMurray, 2017; Kong & Edwards, 2016; Shultz, Francis, & Llanos, 2012) and vowel duration for English tense/lax vowels (Kim & Clayards, 2019). A common method to quantify individual cue weights is to use regression coefficients fit to each individual's responses (e.g., Shultz et al., 2012), or to use by-individual deviations from the population coefficient for a single regression model fit to all data ('random slopes,' e.g., Clayards, 2018; see also Schertz & Claire, 2019 for discussion of different methods). Using these methods, researchers have tried to determine if individuals differ from each other systematically by looking at whether individuals' cue weights are correlated across dimensions. In other words, they asked whether those with larger than average primary cue coefficients have smaller than average or larger than average secondary cue coefficients.

Results have been mixed, with some studies finding a weak or non-significant relationship between cues in perceiving a contrast (Shultz et al., 2012; Clayards, 2018 for F0 and VOT for English stop voicing) or positive correlations (Clayards, 2018; Kim & Clayards, 2019

[1] We use the term 'secondary' to indicate all non-primary cues, not distinguishing between the second most important and any others. Secondary is therefore interchangeable with 'non-primary.'

for vowel formants and duration for English tense-lax vowels). For the Korean three-way laryngeal contrast, one study found positive, negative, or no correlation between cues depending on the contrast pair (Kong & Lee, 2018). Clayards (2018) examined coefficients for individuals across different contrasts in English (e.g., tense/lax vowels and word-final fricative voicing) and found that positive correlations were the most common for both primary and secondary cues (e.g., individuals with larger coefficients for the secondary cue in tense/lax vowels also had larger coefficients for the secondary cue in final fricatives). This suggests that differences between individuals may be systematic and not tied to particular contrasts or dimensions (cf. Hazan & Rosen, 1991). Some researchers have argued that more use of a secondary cue is associated with more gradient sensitivity to primary cues, using a visual analog scaling task rather than a categorical decision task (Kapnoula et al., 2017; Kong & Edwards, 2016). Thus, the positive correlations in Clayards (2018) and the relationship between gradient sensitivity and cue use (Kapnoula et al., 2017; Kong & Edwards, 2016) both point to some listeners' speech perception being more closely tied to the acoustics of the stimulus than others. However, as noted above, not all studies have found this relationship.

Individual variability in secondary cue use may also play an important role in sound change. Some sound changes involve a non-primary cue taking over the role of a primary cue as occurs in 'tonogenesis' (Kingston, 2011), for example, the case of some younger speakers of Afrikaans shifting from VOT to f0 to signal a voicing contrast (Coetzee, Beddor, Shedden, Styler, & Wissing, 2018). In contrast, in the case of Shanghai Wu, a non-primary cue (i.e., breathiness) to the register contrast is losing importance (Gao, 2016). We thus may expect to observe increased individual variability in the use of non-primary cues in a variety undergoing sound change (though see Coetzee et al., 2018 perception data). Conversely, since Shanghai Wu is undergoing a loss of a non-primary cue, rather than an increase in non-primary cue importance, we may see more individual variability in a contrast in a variety that is *not* undergoing the loss of a non-primary cue (Jiashan Wu), and is therefore more dependent on multiple dimensions (e.g., Mayo & Turk, 2005 find larger differences between individuals, in this case adults and children, on contrasts with a larger role for the non-primary cue).

1.2. The perception of voice quality cues

Voice quality, one of the cues of interest in this study, plays different roles in speech perception in different languages. Furthermore, whether voice quality is used to signal a contrast and whether listeners are perceptually sensitive to voice quality are only partially related. This section summarizes the perception of voice quality in three types of languages.

First, in some tone languages, voice quality can be either a phonemic dimension that is independent of pitch, or it can be the main cue to a contrast that also has pitch differences. For example, in Yi, tones with the same pitch can be associated with different phonation types, and listeners rely on phonation cues to distinguish the tones (Kuang, 2011). Jalapa Mazatec (Garellek & Keating, 2011) provides a similar case, where phonation contrasts are independent of pitch contrast, with each tone (low, mid, high) associating with different phonations (laryngealized, modal, breathy). In addition, phonation cues also appear to be perceptually important when different voice qualities are associated with two tones of similar pitch (although not identical). For example, White Hmong has two tones with similar pitch (21 versus 22 in Chao numbers) that also contrast in modal versus breathy voice. Here listeners were found to attend to voice quality information but ignore changes in pitch height or contour (Garellek, Keating, Esposito, & Kreiman, 2013). Similarly, in Sgaw Karen, pitch information is limited and instead, voice quality is crucial for listeners to distinguish the tonal contrast (Brunelle & Finkeldey, 2011). Thus, voice quality can be an independent contrast, or it can be the primary cue in cases where the contrast also includes some difference in pitch.

Second, some languages use phonation as a redundant or non-primary cue associated with certain tones. For example, in Mandarin Chinese, creaky voice facilitates the perception of tone 3, a dipping tone (Kuang, 2013). Similarly, creaky voice in Cantonese tone 4 syllables increases tonal identification accuracy (Yu & Lam, 2011). In Black Miao, tones contrasting in both phonation and pitch are better distinguished than tones contrasting only in pitch (Kuang, 2013). In Northern Vietnamese, creakiness turns out to be as important as pitch in tonal categorization for one pair of tones (Brunelle, 2009).

Third, some languages like English do not use phonation to mark contrasts. Nonetheless, English listeners' perception of talker pitch is influenced by spectral slope such that sounds with tenser/flatter spectral slope are heard as having a higher pitch for both synthetic and resynthesized speech (Kuang & Liberman, 2015; Kuang, Guo, & Liberman, 2016). This may allow listeners to normalize for pitch range (Honorof & Whalen, 2005) by making use of changes in voice quality that occur in the higher parts of speakers' pitch range (Hollien, 1974). However, listeners were not sensitive to the steepness of the spectral slope (as long as it was not entirely flat), i.e., the degree of breathiness/tenseness.

To summarize, voice quality plays a variety of roles cross-linguistically, and is used by listeners for different purposes. The next section gives a brief introduction to the target language of this study, Wu Chinese, in which voice quality facilitates tonal register contrasts.

1.3. Wu Chinese

Wu Chinese is spoken in Shanghai, Zhejiang province and southern Jiangsu province of China. The two target dialects, Shanghai and Jiashan, are both sub-dialects of Wu. According to Yip (1980, 2002), a register feature [+/– Upper] divides the tonal space according to pitch: [+Upper] indicates a higher pitch range (corresponding to the historical *yin* tones) and [–Upper] indicates a lower pitch range (corresponding to the historical *yang* tones) creating a tonal register contrast. Historically, the register was also related to initial consonant voicing, that is, voiceless consonants only occurred in upper registers (*yin* tones) and voiced consonants only occurred in lower registers (*yang* tones). While the restriction on the distribution of consonant voicing and register still exists in non-initial position, the voicing contrast has been lost in initial position (e.g., Chen & Gussenhoven, 2015).

The tone inventory of Shanghai and Jiashan dialect are represented using the register features in **Table 1** together with Chao numbers (Chao, 1930) indicating the pitch contour (numbers 1–5 stand for lower–higher pitch). Tone notations of Shanghai Wu are from Xu and Tang (1988); as there is no previous study on Jiashan Wu, the transcriptions are based on the first author's experience as a native speaker and Yu (1988) on Jiaxing Wu, a highly similar dialect spoken nearby.

There are four types of tones: falling, level, rising, and checked. Checked tones are relatively short in vowel duration and end in a glottal stop. Both dialects have fewer than eight tones due to historical mergers. Shanghai dialect merged the level and the rising tones in the upper register, while falling, level, and rising tones were merged into one rising tone in the lower register (Qian, 1992), leaving a five-tone system. In Jiashan, only the lower-register level and rising tones were merged, resulting in a rising tone (Yu, 1988) and a seven-tone system.

Note that the tone notations in **Table 1** do not accurately match the F0 trajectories, but are rather an abstract representation. Various other notations are proposed by different researchers (as discussed in Chen & Gussenhoven, 2015 for Shanghai Wu), as the pronunciations are varied across individuals and generations. However, no one notation

Table 1: Tone inventory of Shanghai Wu and Jiashan Wu.

Dialect	Register	Falling	Level	Rising	Checked
Shanghai	*Yin*/+Upper (modal)	53		34	55
	Yang/–Upper (breathy)			23	12
Jiashan	*Yin*/+Upper (modal)	53	44	35	55
	Yang/–Upper (breathy)	31		13	12

can accurately reflect the exact F0 trajectories given the large variation. Therefore, similar to differences between formant values and distinct IPA symbols, the tones represented in Chao numbers are phonemic and do not necessarily correspond to the actual phonetic realization. For example, related to the current study, the Jiashan falling tone is realized with a steeper contour in the upper register than the lower register while the Shanghai rising tone is realized with a steeper contour in the lower register than the upper register, although the tone notations do not reflect such differences. For the checked tone, while it is always produced with shorter vowel duration, speakers vary in how audible the glottal stop is.

While pitch range signals the register contrast, it is not the only cue. Others involve voice quality, pitch contour, and duration. Crucially, the upper tonal register is produced with modal voice whereas the lower tonal register is produced with breathy voice (Cao & Maddieson, 1992; Chen, 2010; Gao, Hallé, Honda, Maeda, & Toda, 2011; Zhang & Yan, 2015; Jiang & Kuang, 2016). In addition, in many Wu dialects, the steepness of the contour in contour tones differs in the two registers[2] (e.g., Chen & Gussenhoven, 2015).

While both Shanghai and Jiashan dialects share the same characteristics of register contrast, the Shanghai dialect is argued to be going through a loss of breathiness in the lower register, at least in production (Gao, 2016; Gao et al., 2011). Based on their acoustic and electroglottographic data, Gao and colleagues found that younger speakers used less breathy voice in the lower register compared to older speakers, possibly due to contact with Mandarin Chinese which does not employ breathy voice in tonal contrasts. However, Shanghai Wu listeners do seem to use voice quality in perception despite the change in production. An early study by Ren (1992) found that breathy voice was a perceptual cue for the lower register. A later perception study by Gao, Hallé, and Draxler (2019) found that voice quality did influence perception for both natural (produced by a trained phonetician) and synthesized stimuli. While Zhang and Yan (2015) found that Shanghai listeners heavily relied on F0 information for the register distinction, they also used breathiness as one of the non-primary cues. It is not known how much Shanghai listeners use breathiness relative to other Wu listeners, for whom this cue is not being lost. Direct comparison between listener groups is required to determine this.

To summarize, Shanghai Wu and Jiashan Wu both incorporate a breathy-modal distinction into the tonal register contrast as a non-primary cue. The two dialects have similar phonological systems, although Jiashan has a slightly richer tone inventory and stronger voice quality distinction. Moreover, as Shanghai is in the process of losing breathiness, its listeners may also exhibit some characteristics of listeners of non-phonatorily contrastive languages like English, that is, relying less on breathiness when compared with Jiashan listeners.

[2] Researchers vary in whether they consider the contour difference to be phonetic or phonological. Some treat it as an underlying difference while others consider it different phonetic realizations of the same phonological contrast.

1.4. *Current study*

This paper examines the perceptual difference between dialects and across individuals on the tonal register contrast in Chinese Wu dialects, by manipulating three cues: pitch height (F0 onset), voice quality, and contour (pitch slope). We compare the falling tone pair in Jiashan Wu and the rising tone pair in Shanghai Wu (Experiment 1) and the checked tone pair in both languages (Experiment 2). Because the checked pair is the same in both dialects we can compare both groups of listeners on the same stimuli in Experiment 2. Unlike previous work on perception of voice quality, we explore breathiness as a perceptual cue in fine detail by creating a continuum from natural endpoints that changes in all aspects of the acoustic space for breathiness. Previous studies either only have two levels (breathy versus modal, e.g., Gao et al., 2019; Zhang & Yan, 2015), or manually modified only certain parameters (e.g., Garellek et al., 2013, who modulated H1–H2, H2–H4, H4–2 kHz, and 2 kHz–5 kHz). We also investigate how different dialectal experience caused by sound change is manifested in the perception of multiple cues at both the group level and the individual level by comparing Shanghai and Jiashan Wu listeners on stimuli from both dialects. This paper addresses the three questions raised earlier.

First: Averaging over listeners within each dialect, how important is a non-primary cue (i.e., voice quality) in a multidimensional contrast and does it vary by tone contrast, or is it consistent for different tone pairs? The current study manipulates the three cues independently and includes a five-step breathy-modal continuum. This extends previous studies on cue weighting in Chinese Wu dialects by allowing different combinations of ambiguous cues. We predict that voice quality will significantly influence listeners' perception, at least for Jiashan listeners, but we expect it to be less important than pitch height. The relative importance of voice quality and pitch contour may change depending on the tone pair.

Second: As sound change is taking place in Shanghai, does the decreasing use of the voice quality in production result in different cue weighting than for traditional Wu (i.e., Jiashan) listeners? This study investigates how a dialectal background difference in degree of a non-primary cue production (i.e., voice quality) affects listeners' cue weights. To understand whether listeners from the two dialect groups show different perceptual strategies, and whether the potential difference is caused by the acoustic cues in the stimuli or differences in the listeners, this study examines listeners' cue weightings when they are exposed only to their native dialect (Experiment 1) and when they listen to both dialects (checked tones in Experiment 2). Because the Shanghai dialect is thought to be less breathy, we expect voice quality to have only a small effect on listeners' responses when they listen to the Shanghai stimuli. If both sets of listeners respond to voice quality in the same way, we expect voice quality to have a bigger effect when they listen to Jiashan stimuli, which have a larger voice quality difference. However, if Shanghai listeners are not as sensitive to voice quality as Jiashan listeners we expect their responses to be less affected by voice quality than Jiashan listeners when listening to Jiashan stimuli.

Relevant to the comparison in Experiment 2 is the question of to what extent listeners have experience with the other dialect. It should be noted that both dialects are mutually intelligible, and there are TV shows broadcast in both dialects. Since Shanghai is a large city it may be the case that Jiashan speakers have more opportunities to be exposed to Shanghai Wu than vice-versa. However, it is also the case that Shanghai listeners have opportunities to be exposed to dialects with larger breathy-modal contrasts, although not necessarily Jiashan Wu. Within the city of Shanghai, there are other dialects of Shanghai Wu that differ from the more common 'downtown Shanghai Wu' used in this study in that they are more typical Wu dialects with larger breathy-modal contrast. Moreover, older Shanghai speakers produce breathier lower register words than younger speakers (Gao, 2016) and residents of Shanghai could also be exposed to speakers of other, mutually-intelligible Wu dialects with a greater breathy-modal contrast. Given these factors it is difficult to say whether one dialect group is more familiar with the other dialect or not.

Third: Do individuals show structured differences in cue ordering or cue magnitude, or are differences between individuals random variation? Does the structure of individual variability differ by dialect? To address these issues, this study examines individual variability in relative cue weighting. Previous work on individual variability has examined only two cues to a contrast. By examining three cues we are able to better observe how cue weights relate to each other. Group-level results are sometimes insufficient to understand how a multidimensional contrast is perceived, as the results are averaged over all participants, giving only one pattern of cue weighting. How individual cue weights diverge from the average pattern is important for fully understanding the phonological contrast of the target language as well as the perceptual system of individuals. Individual variability in cue weights may be completely random, or structured (correlated). Individual differences could be small, with individuals only differing in the magnitude of the cues, while sharing the same ordering of cue weights, or larger, with individuals differing in the ordering of cue importance. Because so little is known about individual variability in cue use in perception, especially with more than two cues, it is difficult to make a priori predictions. To examine individual differences, we use the correlation matrix of the random effects of a mixed-effects model fit to the data. A methodological contribution of this paper is showing how this component of mixed-effects models, which is typically ignored in phonetic studies, can be used to understand the structure of individual differences.

2. Methods

The current study consists of two experiments. The first experiment examined Jiashan and Shanghai listeners' use of cues in stimuli from their own dialects; the second experiment again evaluated the three acoustic cues but exposed listeners to stimuli from both dialects and examined whether listeners have different cue weightings for the two sets of stimuli.

2.1. Participants

Two groups of listeners participated in the study. Thirty-four native Jiashan Wu speakers (5 males, 29 females, aged 19–62 with mean of 34) were recruited in Jiashan and 35 native Shanghai Wu speakers (11 males, 24 females, aged 18–56 with mean of 22) were recruited in Shanghai (34) and Montreal (1). No participant reported hearing loss.

2.2. Stimuli

2.2.1. Experiment 1

In the first experiment, the stimuli varied by five steps in both pitch height and voice quality within the natural range between the upper and lower register, as determined by natural productions of native speakers. Pitch contour, however, only had two levels: the upper register contour and the lower register contour (as in Kirby, 2014 on Khmer). The number of stimuli were limited in this way so that the experiment could be conducted in a reasonable amount of time, while allowing the cue of primary interest (i.e., breathiness) to vary on a continuum. Moreover, we did not expect the pitch contour to play a large role in listeners' perception. This is because various Chinese Wu dialects do not contrast degree of steepness (Chao, 1928; Qian, 1992) and while some dialects do exhibit different degrees of steepness of the two contours, many researchers treat this as different phonetic realizations of the same underlying form due to articulatory constraints.

The endpoint sounds were selected from recordings of a previous production study on several Wu dialects (Jiang & Kuang, 2016). Various acoustic parameters of breathy voice (*H1, *H1–*H2, *H1–*A1, *H1–*A2, *H1–*A3, CPP) were measured in all 12 speakers.

The speaker that produced the largest breathy-modal contrast in each dialect was selected, based on a combination of auditory saliency of breathiness perceived by native speakers and extremity of measures (which largely coincided). From the productions of these two speakers the minimal pairs that had the greatest breathy-modal distinction were chosen. The original productions (i.e., two endpoints) of the Jiashan stimuli were a pair of upper and lower tonal register words ([ka 53] 街 'street' and [ka 31] 扛 'carry on the shoulder') with falling tonal contour produced by a female native speaker aged 43. The two monosyllabic words varied in three dimensions: voice quality (the upper register word [ka 53] is modal while the lower register word [ka 31] is breathy), pitch height ([ka 53] has higher pitch), and pitch contour ([ka 53] is steeper). The two endpoint productions for the Shanghai stimuli were produced by a female native speaker aged 24. The syllables were also ([ka]) while the tones were rising ([ka 23] 茄 'eggplant' and [ka 34] 价 'price'). Rising tones were avoided in Jiashan because they are realized more like dipping tones, sometimes including creakiness at low pitch in a low register word, which undermines the breathy voice quality examined in this study. Falling tones were avoided in Shanghai Wu because there is no low register falling tone (see **Table 1**).

The natural recordings were then modified using a combination of TANDEM STRAIGHT (Kawahara et al., 2008) and the PSOLA method in Praat (Boersma & Weenink, 2016) to create a series of stimuli. TANDEM STRAIGHT provides high quality source-filter resynthesis and allows the user to interpolate between two natural recordings. It morphs a pair of sounds in all dimensions (e.g., F0, duration, spectrum) simultaneously and holistically without reference to specific dimensions. The resulting continuum has the same recording quality and naturalness as the originals and varies in all dimensions of the original two sounds. TANDEM STRAIGHT has been successfully used in studies of many different phonological contrasts cross-linguistically (e.g., fricative place in German and English, nasal place in Japanese, vowel contrasts and sung melodies in Japanese: Bukmaier, Harrington, & Kleber, 2014; McAuliffe & Babel, 2016; Sadakata & Sekiyama, 2011; Yonezawa, Suzuki, Mase, & Kogure, 2005). We believe that this method is also appropriate for creating a natural breathiness continuum that varies in all acoustic dimensions of breathiness, providing a more thorough modification than previous studies that only focus on a few dimensions.

To create the stimuli, we first normalized vowel duration using the PSOLA method between the two registers within each dialect (duration of voiced portion: JS 219 ms; SH 283 ms). For the Jiashan stimuli, the VOT of the initial consonant /k/ was similar for the two sounds, so there was no need to normalize. For the Shanghai stimuli, we normalized VOT, although both were very close to 19 ms. Second, we created a second copy of each word that had the pitch contour (but not pitch height) of its pair. To do this, we extracted the pitch contour from both sounds. The upper contour was lowered by 100 Hz (the difference between the two sounds) by PSOLA, and was then superimposed on the lower register word. The lower contour was shifted up by 100 Hz and was superimposed on the upper register word. Third, we used the two pairs of upper and lower register words matching in contour shape (one natural, one manipulated) to create two five-step continua in TANDEM STRAIGHT. Having controlled the contour, the two continua varied in both pitch height and voice quality, one with a flatter contour and the other with a steeper contour. Since STRAIGHT does not generate an equal-stepped continuum, we set the program to generate 15 stimuli, from which we picked five that varied in pitch height in 25 Hz steps. Since the manipulation changes all aspects of the source, the breathiness also varied in these five steps.[3] Fourth, we created four new versions of each stimulus varying

[3] Note that the generated stimuli are not precisely equal-stepped due to the noise introduced in STRAIGHT continuum generation, although this is unavoidable and we made sure that they are reasonably equal-stepped, given the large range of the register contrast.

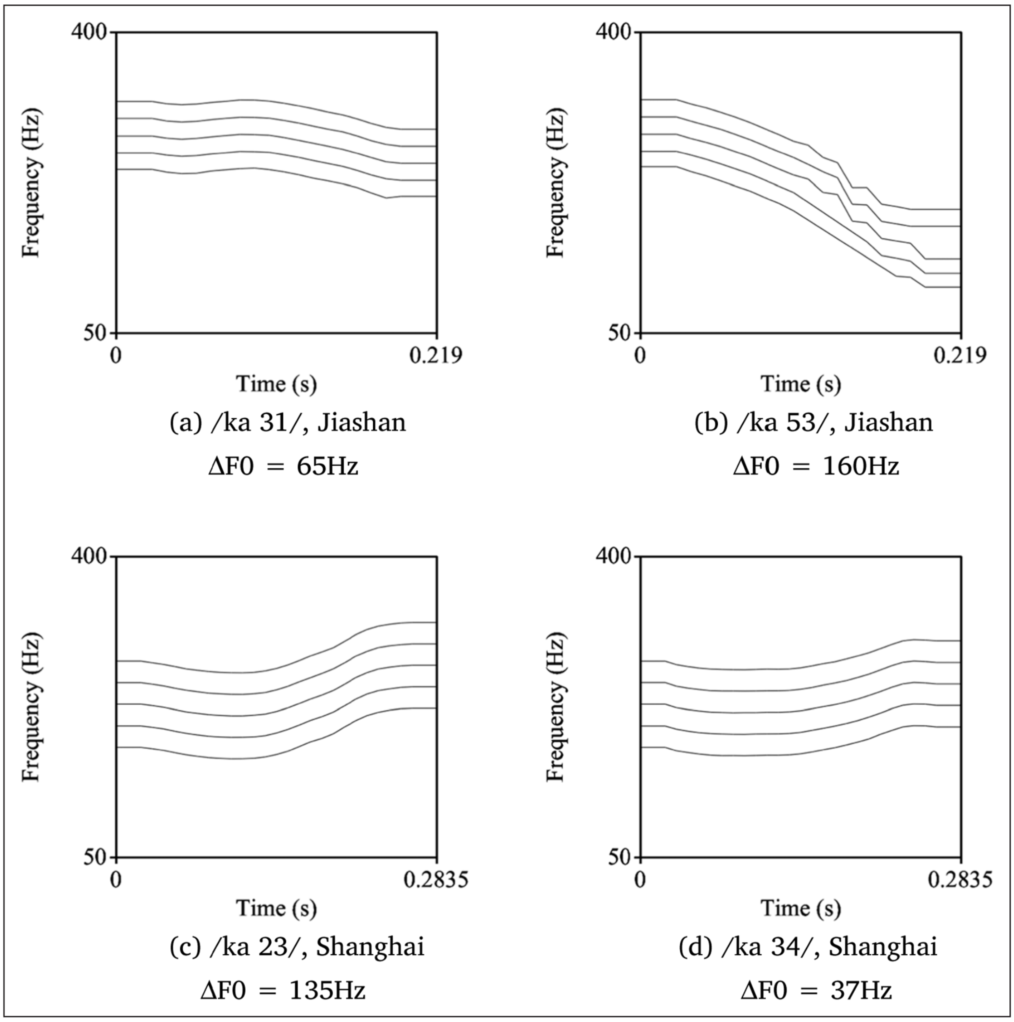

Figure 1: Pitch contours (left panels lower register, and right panels upper register) and pitch height (five steps) used in the two dialects. ΔF0 refers to the F0 change between the onset and the offset of the syllable. The breathiness continua also have five steps, but are not shown schematically.

in pitch height. We manipulated the pitch height of each stimulus by shifting it upwards or downwards in 25 Hz steps to create five steps total (Jiashan stimuli: F0 onset from 240 Hz to 340 Hz; Shanghai stimuli: from 190 Hz to 290 Hz). Together this process resulted in two 5 × 5 continua for each dialect varying in breathiness and pitch height, one with the flatter contour and one with the steeper contour for a total of 50 stimuli (5 × 5 × 2) per dialect. **Figure 1** shows the 10 different pitch contours used for each of the two dialects.

2.2.2. Experiment 2

In the second experiment, the stimuli included three five-step continua, one for each cue, and for each continuum the other two cues were held at the acoustic mid-point between the two registers. This aimed to examine each cue independently, and by holding the other cues at what is expected to be the most ambiguous point, it minimized the reliance on other cues to best show listeners' use of the target cue.

The second experiment used a different pair of tones (i.e., checked tone) where the two dialects share the same phonological representation of the words used (i.e., [ka 55] 夹 'clip' and [ka 12] 挤 'jostle') and are mutually intelligible. Listeners heard both sets of stimuli from the two dialects, in order to examine whether the two groups of listeners differ when exposed to the same stimuli.

The original productions were a pair of /ka/ syllables with the upper and lower register checked tones. Recordings from the same speakers as in Experiment 1 were used to create three new continua respectively.

As in Experiment 1, stimulus construction was the same for both dialects and used a combination of TANDEM STRAIGHT and PSOLA in Praat. Prior to creating the continua, the two endpoint productions were normalized in duration (JS: 122 ms; SH: 120 ms) and VOT within dialect. We did not normalize the F0 range between the two dialects, because shifting the F0 could result in a change in breathiness, especially when the change was large.[4]

BREATHNESS CONTINUUM: As before, we used TANDEM STRAIGHT to create a 15-step continuum from the normalized endpoints. From this we picked five steps that were equally spaced in F0 onset (25 Hz steps, Shanghai 190 Hz to 290 Hz; Jiashan 240 Hz to 340 Hz). We used Step 3 from this five-step continuum as the acoustically ambiguous step in terms of breathiness for the other two continua. CONTOUR CONTINUUM: Using Praat, we extracted the endpoint F0 contours from the normalized endpoints and shifted them such that the midpoint F0 of both was midway between the originals. This created two F0 contours that only differ in the contour shape, which were used as the two endpoints of the contour continuum (see **Table 2** for all values of contour stimuli). We

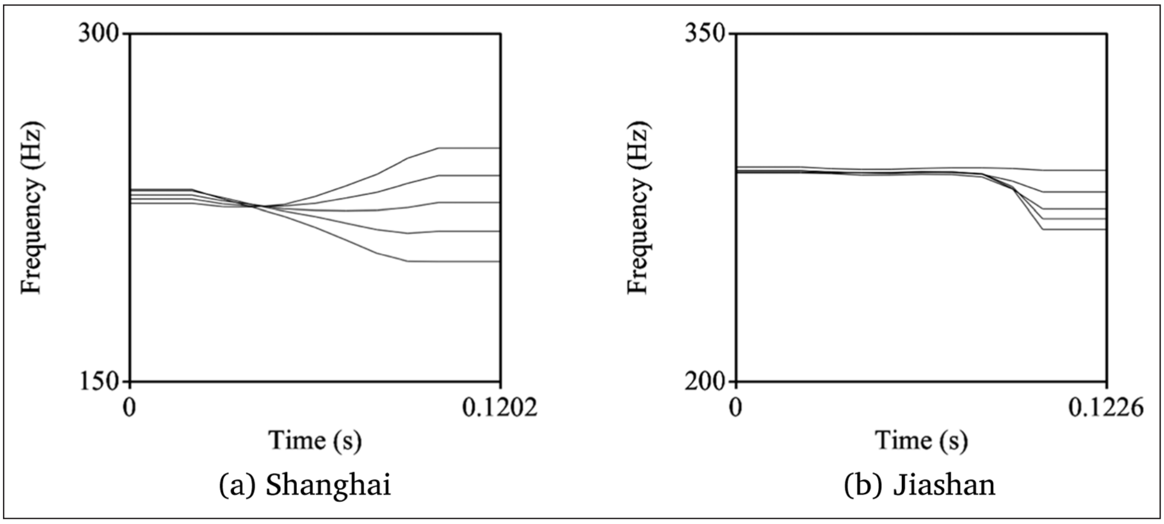

Figure 2: Contour continua of Shanghai (left) and Jiashan (right) used in Experiment 2.

Table 2: Onset and offset f0 (Hz) for each step of the contour continua of Experiment 2.

Contour	Shanghai talker		Jiashan talker	
Step	Start f0 (Hz)	End f0 (Hz)	Start f0 (Hz)	End f0 (Hz)
1	230	253	291	291
2	232	245	291	279
3	234	232	291	268
4	236	220	290	262
5	237	210	290	253

[4] Test stimuli showed that manipulating F0 affects the amplitude of the harmonics to some extent, which affects the degree of breathiness. Since the continua already have a large F0 range (100 Hz), we decide not to further introduce more confounding factors.

then used a Praat script to linearly interpolate between the two endpoints to create a five-step continuum between these endpoints. We used Step 3 of the contour continuum as the acoustically most ambiguous step to be used for the pitch height and breathiness continua. PITCH HEIGHT CONTINUUM: We chose five F0 onset values to match those chosen for the breathiness stimuli (25 Hz steps, Shanghai 190 Hz to 290 Hz; Jiashan 240 Hz to 340 Hz).

The final contour continuum had the five contours applied to Step 3 of the breathiness continuum with the pitch onset value (i.e., pitch height) from Step 3 of the pitch continuum. The final breathiness continuum had Step 3 of the contour continuum and the pitch onset value from Step 3 of the pitch continuum applied to each of the five steps of the breathiness continuum. The final pitch continuum used Step 3 of the breathiness continuum and the contour from Step 3 of the contour continuum and varied in pitch onset according to the five steps.

The above manipulation created three five-step continua of pitch height, breathiness, and contour while holding the other two factors constant and acoustically ambiguous. Each stimulus was repeated five times, resulting in 150 trials in total (5 steps × 3 continua × 5 repetitions × 2 dialects = 150).

All stimuli used in the two experiments can be found at https://osf.io/u7er5/, together with corresponding acoustic measures for voice quality (e.g., H1–H2, H1–A1) and plots showing how these measures change along the five-step voice quality continuum. By all measures, the Jiashan stimuli show a larger breathy-modal contrast than the Shanghai stimuli for both experiments, supporting our claim that the TANDEM STRAIGHT method created breathiness continua that vary simultaneously in all acoustic measures.

2.3. Procedures

The experiments were conducted using Matlab in a quiet room. All participants took part in both experiments with a five-to-ten minute break in between. They first listened to 250 sounds of their own dialect (i.e., Experiment 1) and then 150 sounds, 75 from each dialect (i.e., Experiment 2). All stimuli were played in random order. On each trial, participants heard a single syllable and the display presented two Chinese characters corresponding to the upper and lower register words together with the numbers "1" and "2" corresponding to the associated keys. The association between the number and the character varied randomly across the whole experiment. Listeners pressed the key associated to the word they perceived. After selecting the answer, they pressed the space key to proceed to the next trial. In Experiment 2, participants were told that they may hear more than one speaker, but no information about speaker or dialect was given.

3. Results

In this section, we first provide group-level results of the two experiments, showing how each of the two listener groups makes use of the three cues on average. We then present results for individual variability and show how listeners differ in a structured manner.

3.1. Experiment 1: Group results

Experiment 1 examines listeners' perception of tonal register contrast (upper versus lower) of their native dialect when exposed to stimuli with different degrees of breathiness and pitch and with two contours. In order to investigate the relative importance of the three cues, two mixed-effects logistic model were fit for the two dialect groups respectively to model participants' response as a function of the three variables.

The data were fitted with a Bayesian mixed-effects logistic regression (MELR) model in R using version 1.0-4 of the blme package (Chung, Rabe-Hesketh, Dorie, Gelman, &

Liu, 2013) using the default settings for priors. One advantage of a Bayesian MELR over the more widely used non-Bayesian MELR (e.g., using lme4 in R) is the ability to fit a maximal random-effect structure (all possible slopes and correlation terms; Barr, Levy, Scheepers, & Tily, 2013) without convergence issues which frequently arise for non-Bayesian methods (including when applied to our data), usually related to 'impossible' 1/–1 correlation values (Nicenboim & Vasishth, 2016; see Vasishth et al., 2018 for a tutorial). The dependent variable was the register response (upper = 1, lower = 0). All predictors were centered and standardized by subtracting the mean and dividing by two standard deviations. This makes the coefficients comparable (e.g., in **Table 3**, the coefficient estimates of pitch are the highest among the three cues, indicating that it is the primary cue) and minimizes collinearity between main effects and interactions (Gelman & Hill, 2006). More importantly, centering means the interpretation of a certain main effect is averaged over other variables this specific variable interacts with. The fixed-effect predictors were voice quality step *breathiness* (coding for five steps are scaled in range [–0.7, 0.7]), pitch height step *pitch* (coding for five steps were scaled in range [–0.7, 0.7]) and pitch contour step *contour* (two levels: upper = 0.5 lower = –0.5). Two-way and three-way interactions were also included because changes in one cue may have an effect on participants' perception of the other cues. The random effects included by-participant random intercepts and random slopes for breathiness, pitch, and contour and their interactions, to account for individual differences in cue weights, as well as all possible correlation terms (the *maximal* structure), to measure potential correlations in cue usage among individuals.[5]

3.1.1. Cue weights in Shanghai Wu

Figure 3 shows the mean responses from the Shanghai participants. **Table 3** shows the model results for fixed-effects terms. All three cues had a significant effect, and only the pitch-breathiness term was significant among all interactions. Note that since all variables are centered, the main-effect coefficient estimates indicate how much the log-odds ratio of responding with the upper register changes as the variable shifts by one unit, averaging over other factors (e.g., the β coefficient of contour means the effect of contour at the average pitch and breathiness values).

According to the coefficient estimates, pitch was the most important cue ($\beta = 6.67$, $SE = 0.36$, $p < 0.001$), much stronger than contour ($\beta = 1.26$, $SE = 0.18$, $p < 0.001$),

Table 3: Summary of the fixed effects for the model of Shanghai participant data from Experiment 1.

	Estimate	*SE*	*z*-value	Pr(>|z|)
Intercept	0.59	0.16	3.54	**<0.001**
breath	0.55	0.11	4.83	**<0.001**
pitch	6.67	0.36	18.76	**<0.001**
contour	1.26	0.18	6.86	**<0.001**
breath × pitch	0.74	0.32	2.25	**0.024**
breath × contour	0.23	0.19	1.16	0.244
pitch × contour	0.33	0.39	0.86	0.386
breath × pitch × contour	1.43	0.73	1.94	0.052

[5] Model syntax: bglmer (response ~ BreathDegree.std * PitchDegree.std * Contour.std + (1 + BreathDegree.std * PitchDegree.std * Contour.std|participant), data = df, family = "binomial", control = glmerControl(optimizer = "bobyqa", optCtrl = list (maxfun = 100000))).

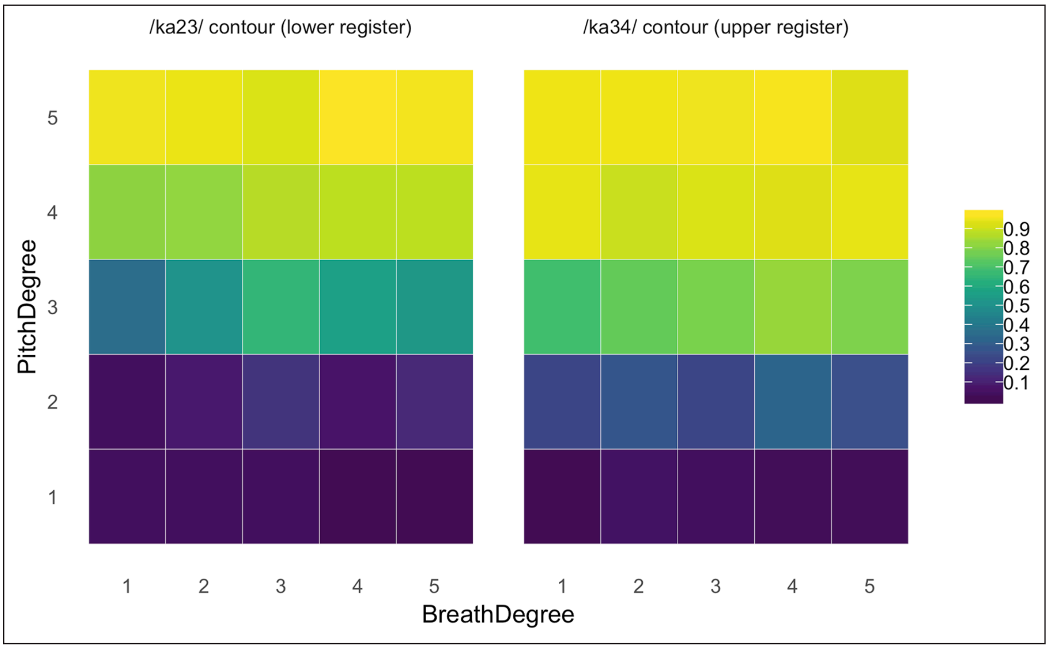

Figure 3: Percentage of upper register response from Shanghai participants. The x axis is the breathiness continuum (1 = breathy, 5 = modal). The y axis is the pitch continuum (1 = low pitch, 5 = high pitch).

while breathiness was the least important ($\beta = 0.55$, $SE = 0.11$, $p < 0.001$).[6] Moreover, the two-way interaction between breathiness and pitch ($\beta = 0.74$, $SE = 0.32$, $p = 0.024$) shows that the higher the pitch, the more important voice quality is. There was a trend for a three-way interaction between the cues ($\beta = 1.43$, $SE = 0.72$, $p = 0.052$), indicating that how much the effect of breathiness is affected by pitch is itself modulated by the specific contour. Together these interactions suggest that the higher pitch onsets and the upper register contour were slightly more ambiguous than the lower pitch onsets and lower register contour.

Figure 3 shows the corresponding heat map of the participants' categorization. The left panel represents the results for the lower register tone contour and the right panel represents the results for the upper register contour. The lighter color indicates higher percentage of upper register responses. In terms of cue weighting, compatible with the model predictions, there was indeed a large effect of pitch, as indicated by the change of colors along the y axis in both panels. Effects of breathiness and contour are much weaker, and are most visible in regions with ambiguous values of pitch (step 3). The effect of breathiness can be seen in a color change along the x-axis and the effect of contour can be seen in lighter colors in the right panel, consistent with model predictions.

3.1.2. Cue weights in Jiashan Wu

The Jiashan participants, however, show a different pattern. As displayed in **Table 4**, the Jiashan model found that contour had the biggest effect ($\beta = 3.08$, $SE = 0.46$, $p < 0.001$) followed by pitch ($\beta = 1.63$, $SE = 0.30$, $p < 0.001$) and breathiness ($\beta = 0.86$, $SE = 0.16$, $p < 0.001$).[7]

[6] Likelihood ratio tests of the contribution of each cue, removing all pitch/breathiness/contour terms, give the same ordering: Pitch ($\chi^2 = 6462.3$, $p < 0.001$), Contour ($\chi^2 = 304.65$, $p < 0.001$), Breathiness ($\chi^2 = 62.31$, $p < 0.001$). Note that the model without pitch terms did not converge, presumably because the most explanatory variable for the data was omitted, so the Likelihood ratio test result for pitch is approximate.

[7] Likelihood ratio tests of the contribution of each cue, removing all pitch/breathiness/contour terms, give the same ordering: Contour ($\chi^2 = 3342.2$, $p < 0.001$), pitch ($\chi^2 = 1035.8$, $p < 0.001$), breathiness ($\chi^2 = 356.7$, $p < 0.001$).

Table 4: Summary of the fixed effects for the model of Jiashan participant data from Experiment 1.

	Estimate	*SE*	*z*-value	Pr(>|z|)
Intercept	0.73	0.11	6.51	**<0.001**
breath	0.86	0.16	5.14	**<0.001**
pitch	1.63	0.30	5.45	**<0.001**
contour	3.08	0.46	6.62	**<0.001**
breath × pitch	−0.69	0.20	−3.37	**<0.001**
breath × contour	−0.52	0.23	−2.22	**0.02**
pitch × contour	−0.74	0.24	−3.06	**0.002**
breath × pitch × contour	−0.30	0.39	−0.76	0.44

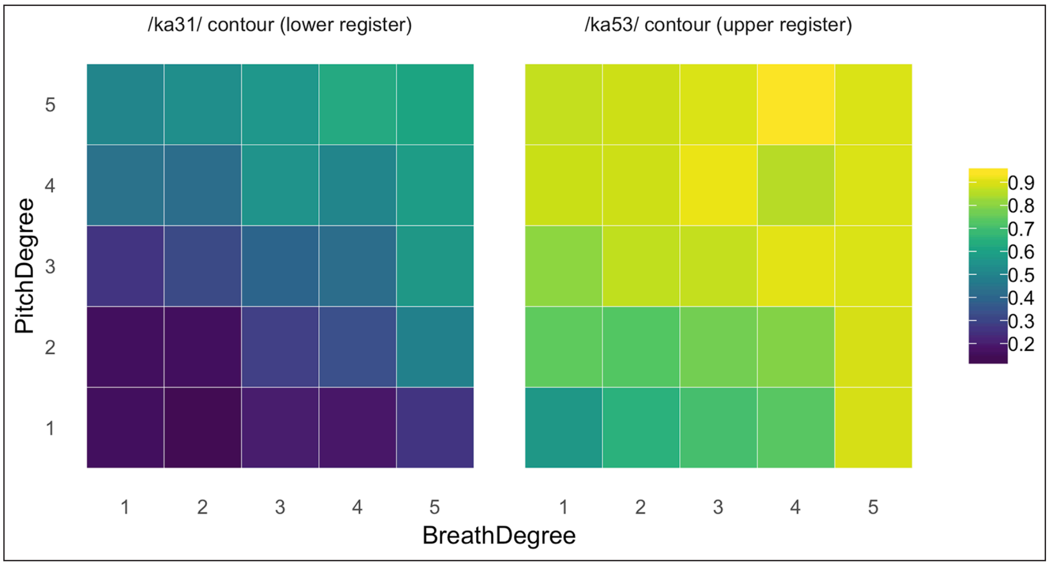

Figure 4: Percentage of upper register response from Jiashan participants (axes as in Figure 3).

Heat maps of the empirical data also reflect this pattern (**Figure 4**). It is obvious that contour has a large effect on the perception of tonal register for Jiashan listeners. When listening to the upper register contour (right panel), listeners consistently hear it as an upper register word. The heat map also shows that pitch and breathiness affect categorization, more so for the lower register contour, as captured by the significant interaction terms between breathiness and contour (β = –0.35, *SE* = 0.18, *p* = 0.04) and between pitch and contour (β = –0.62, *SE* = 0.19, *p* < 0.001), which predict lower breathiness and pitch effects for the upper register contour.

The significant breathiness by pitch interaction (β = –0.69, *SE* = 0.2, *p* < 0.001) suggests that the effect of breathiness is smaller at higher pitch than at lower pitch. The right panel reflects this prediction in that there is a larger change in color at step 1 of the pitch continuum (the bottom row) than at step 5. As with the Shanghai responses, these interactions likely reflect the fact that the continua were not symmetric in how ambiguous the steps were. In particular, the lower register contour seems to have been more ambiguous than the upper register contour.

3.1.3. Summary of cue weights in both dialects

In summary, both groups of participants use all three cues, although both treat breathiness to be least important. Pitch appears to be more important than contour for Shanghai listeners, while contour is the primary cue for Jiashan listeners. In addition, for Jiashan listeners, breathiness and pitch have a smaller effect with the upper register contour; the effect of breathiness is also smaller at higher pitches for Jiashan listeners but larger at higher pitches for Shanghai listeners.

It is worth noting that although the cue weighting only reflects listeners' perception of the stimuli used in this task, the stimuli for the two dialects differ in ways that are representative of the communities speaking the dialect. Moreover, in terms of the manipulation of the three cues, while pitch and contour share the same range between the two endpoints for the two dialects, the range of breathy-modal distinction is smaller in the Shanghai stimuli (see visualizations in the supplemental materials). As a result, there also exists a difference in the degree that the breathiness cues were manipulated between the two dialects. It should also be noted that the two pairs of tones used in Experiment 1 are confounded in contour (falling for Jiashan, rising for Shanghai), which further limits the interpretation of difference in cue weighting. We next present Experiment 2, which investigated the dialectal difference by having all participants listen to the same stimuli, in contrast to Experiment 1.

3.2. Experiment 2: Group results

This section extends the previous experiment by directly comparing the two dialect groups. In this experiment, listeners listen to the same stimuli (i.e., checked tone) produced by talkers from both dialects. Note that checked tones are marked with much shorter syllable duration and overall flatter contour (as shown in **Figure 2** and **Table 2**), which may reduce the use of the contour cue. Checked tones are used for Experiment 2, however, because they are the only pair of tones sharing the same phonological representation between the two dialects (see **Table 1**). When exposed to stimuli differing in the degree of phonation contrast, we expect listeners who are sensitive to breathiness to show smaller breathiness cue weights for Shanghai stimuli with a weak breathy-modal distinction, than for Jiashan stimuli which differ more in breathiness.

Experiment 2 also clarifies one remaining problem in Experiment 1. Specifically, some participants relied almost solely on contour (see Section 3.3 on individual differences below). One possibility is that the effects of pitch and breathiness were masked by a dominant contour cue and weren't observed because of the lack of an intermediate/ambiguous contour value. By breaking down the contour cue into a five-step continuum (while holding the other two cues at mid-point), Experiment 2 better examines how these listeners use all three cues.

The data were again fitted with a Bayesian mixed-effects logistic regression model in R using the blme package (Chung et al., 2013). As with Experiment 1, the dependent variable was the register response (upper = 1, lower = 0). Independent variables were breathiness, pitch, and contour (continuum step as a numerical variable from 1 to 5) and their interactions with talker dialect and participant dialect (for both talker and participant dialect, Jiashan was coded as 0 and Shanghai was coded as 1) as the effect of the cues may vary by talker and participant dialect. Random effects include by-participant random intercepts, and random slopes for the talker dialect and the three cues. Correlations among all random slopes were included. All predictors were centered and standardized by subtracting the mean and dividing by two standard deviations, as in Model 1 (making Jiashan = –0.5 and Shanghai = 0.5 for the talker and participant dialect variables).[8] **Table 5** shows the fixed effects of the model.

[8] Model syntax: bglmer (response ~ (Breath.std + Contour.std + Pitch.std)*dialect.std*part_dia.std + (1 + (Breath.std + Contour.std + Pitch.std)* dialect.std|participant), data = df.rs, family = "binomial", control = glmerControl (optimizer = "bobyqa", optCtrl = list(maxfun = 100000))).

Table 5: Summary of the fixed effects of the model.

	Estimate	*SE*	*z*-value	Pr(>|z|)
Intercept	0.48	0.11	4.28	**<0.001**
Breath	0.43	0.00	4.74	**<0.001**
Contour	0.35	0.08	4.19	**<0.001**
Pitch	3.14	0.19	15.92	**<0.001**
talkerDialect	–4.34	0.27	–15.73	**<0.001**
listenerDialect	1.15	0.22	5.26	**<0.001**
Breath × talkerDialect	0.24	0.17	1.38	0.165
Contour × talkerDialect	0.63	0.18	3.37	**<0.001**
Pitch × talkerDialect	0.86	0.22	3.82	**<0.001**
Breath × listenerDialect	–0.43	0.16	–2.62	**0.008**
Contour × listenerDialect	–0.25	0.14	–1.67	0.091
Pitch × listenerDialect	–0.87	0.38	–2.26	**0.023**
talkerDialect × listenerDialect	–0.18	0.54	–0.34	0.720
Breath × talkerDialect × listenerDialect	0.97	0.31	3.05	**0.002**
Contour × talkerDialect × listenerDialect	–0.13	0.34	–0.39	0.69
Pitch × talkerDialect × listenerDialect	0.42	0.41	1.02	0.30

The results show that all the main effects are significant. This means that all three cues, although redundant, are important in perceiving the tonal register contrast for this tone pair. Secondly, responses depended on listener dialect and talker dialect. Specifically, listeners tended to perceive more lower register words when they listened to a Shanghai talker (indicated by the negative talkerDialect coefficient) probably because the talker had an overall lower pitch range. Furthermore, Shanghai listeners tended to hear more upper register words (indicated by the positive listenerDialeect coefficient). However, of main interest were the interactions between the cues and the talker and listener variables.

We found that talker dialect influenced use of contour ($\beta = 0.63$, $SE = 0.18$, $p < 0.001$) and pitch ($\beta = 0.86$, $SE = 0.22$, $p < 0.001$), indicating that listeners were more influenced by both contour and pitch when listening to the Shanghai talker. Listener dialect interacted with breathiness ($\beta = –0.43$, $SE = 0.16$, $p = 0.008$) and pitch ($\beta = –0.87$, $SE = 0.38$, $p = 0.023$), which reveals that Shanghai listeners were less influenced by breathiness and pitch. Moreover, there was a significant three-way interaction between talker dialect, listener dialect, and breathiness ($\beta = 0.97$, $SE = 0.31$ $p = 0.002$) indicating that listeners were more influenced by breathiness when listening to their own dialect than when listening to the non-native dialect, at least for Shanghai listeners (see **Figure 5**).

Given that the model contains many two-way and three-way interactions, and the cue weighting of specific listener-talker combinations require taking these interactions into consideration, it is useful to calculate the coefficient estimates for usage of each cue as a function of talker and listener dialects. **Table 6** below (also visualized in **Figure 5**) shows the coefficient estimates (with standard errors) of the three cues for different talker-listener combinations, calculated using the *emmeans* package (Lenth, Love, & Hervé, 2017) in R.

Pitch is the primary cue for all talker/listener dialect pairs, being much more important than the secondary cues (breath, contour). This suggests that when listening to checked tones, both Shanghai and Jiashan listeners mainly use pitch to distinguish the two

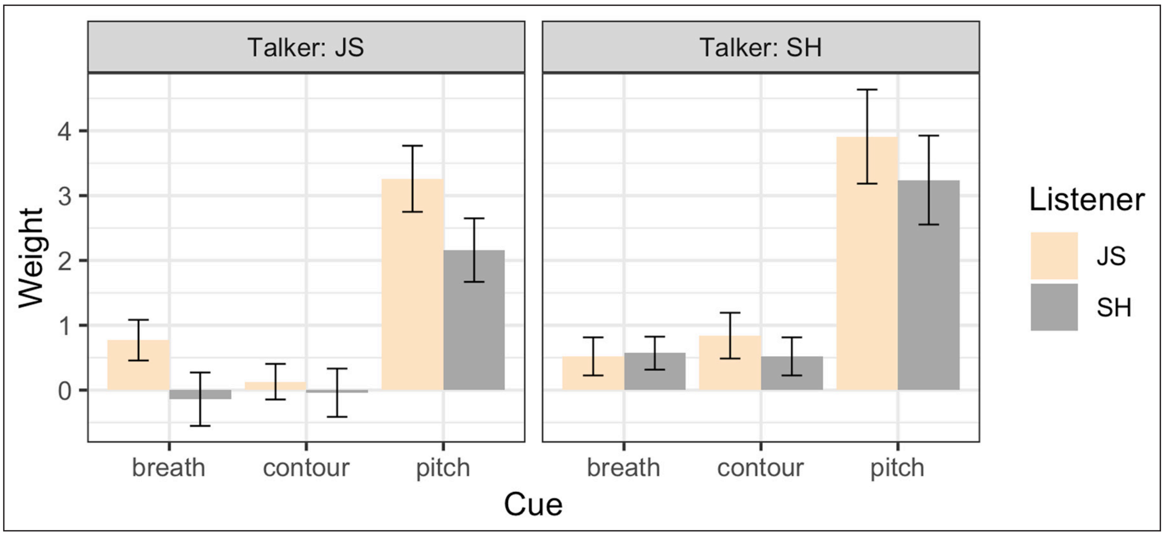

Figure 5: Cue weights for Jiashan (JS) and Shanghai (SH) talker by listener with error bars showing 95% speaker variability intervals, capturing individual variation in cue weights. Experiment 2.

Table 6: Cue weights (coefficient estimates) of the three cues for different talker-listener combinations. *SE*: standard error of coefficient. Experiment 2.

Talker	Listener	Pitch	*SE*	Breath	*SE*	Contour	*SE*	Cue weight
JS	JS	3.26	0.26	0.77	0.16	0.13	0.14	P > B > C
JS	SH	2.16	0.25	−0.14	0.21	−0.04	0.19	P > B ≥ C
SH	JS	3.91	0.37	0.52	0.15	0.84	0.18	P > C > B
SH	SH	3.24	0.35	0.57	0.13	0.52	0.15	P > B ≥ C

registers. This is in contrast to Experiment 1 in which only Shanghai listeners used pitch as the primary cue. Secondly, when listening to the Jiashan talker, the Jiashan listeners had on average a positive cue weight for breathiness while the Shanghai listeners had on average a coefficient close to zero. Thus even when listening to the same stimuli, Jiashan listeners' responses are more affected by breathiness. Finally, when listening to the Shanghai talker, Jiashan listeners had a larger coefficient for contour (their secondary cue for that contrast) than the Shanghai listeners. Thus, regardless of the stimuli, Jiashan listeners seem to have larger secondary cue coefficients. In fact, comparing within stimuli and dimensions (e.g., Jiashan talker pitch) in almost every case the Jiashan listeners have larger coefficients than the Shanghai listeners for the same stimuli and dimension.

Furthermore, Jiashan listeners appear to rely more on breathiness than Shanghai listeners as indicated by the change of cue weighting for different talker dialects. When exposed to a larger breathy-modal contrast (Jiashan Wu talker), Jiashan listeners indeed had a larger coefficient for breathiness than they did when listening to Shanghai Wu. Shanghai listeners on the other hand, had a smaller coefficient for breathiness when listening to Jiashan Wu talker, the opposite tendency to what one would expect given the greater role of breathiness in Jiashan Wu in production. Thus, Shanghai listeners did not use breathiness when they were actually exposed to a larger breathy-modal contrast, while they managed to better perceive such a contrast in their own dialect with a smaller difference.

To summarize, Jiashan and Shanghai listeners share a similar order of cue preference for both talkers, with pitch always being the primary cue. However, Jiashan listeners' coefficients more closely relate to the amount of breathy-modal contrast in the stimuli than Shanghai listeners. Thirdly, both listeners show greater use of voice quality cue when listening to their own dialect.

3.3. Individual variability

3.3.1. Experiment 1

Group-level results found in Experiment 1 indicate that while all three cues significantly affect both groups of listeners' tonal register categorization, Shanghai listeners' perception was more restricted to a single cue (pitch) while Jiashan listeners on average had a more balanced cue weighting. What is not clear from the group results is whether these average patterns were an equally good description of individuals in the two groups. In the Introduction, we hypothesized that the ongoing loss of breathiness in Shanghai Wu might lead to inter-listener variability. On the other hand, having a contrast that depends on more cues might lead to greater listener variability in the Jiashan listeners. In order to test the two hypostheses, we compare variability in individual cue weights between the two listener groups. In addition, we wanted to test if there are patterns across cues in how individuals use them. In other words, does having a large primary cue weight predict having a large secondary cue weight, or does it predict having a small secondary cue weight? Examining the structure of the variability between cues allows us to test these competing hypotheses.

The random effects fit in the models in the previous section, summarized in **Tables 7** and **8**, allow us to examine individual variability in both the ways detailed above. Specifically, random effects are the part of the statistical model that captures the variability within

Table 7: Summary of random effects and correlations of Shanghai participants, Experiment 1. Correlations between interaction terms are not shown.

Name	**Variance**	***SD***	**Correlation**		
Intercept	0.91	0.95			
Breathiness	0.17	0.41	0.50		
Pitch	3.91	1.97	0.61	0.83	
Contour	0.86	0.93	0.54	0.72	0.63
Breath × Pitch	0.61	0.78			
Breath × Contour	0.21	0.45			
Pitch × Contour	1.89	1.37			
Breath × Pitch × Contour	6.42	2.53			

Table 8: Summary of random effects and correlations of Jiashan participants, Experiment 1. Correlations between interaction terms are not shown.

Name	**Variance**	***SD***	**Correlation**		
Intercept	0.34	0.59			
Breathiness	0.68	0.82	0.25		
Pitch	2.71	1.64	0.29	0.93	
Contour	6.86	2.61	−0.23	−0.12	−0.15
Breath × Pitch	0.46	0.67			
Breath × Contour	0.87	0.93			
Pitch × Contour	0.89	0.94			
Breath × Pitch × Contour	1.28	1.13			

groups (in this case, the variability of cue weights among different participants). We can examine the standard deviation in estimates for each cue (*SD* column), which capture the degree of variability in cue weights across participants, as well as the correlations between cues (across participants).

Figure 6 summarizes the variability across individuals in the weight of each of the three cues, for each listener group. For each cue, the fixed-effect estimate from the model plus or minus twice the random effect standard deviation gives the range of cue values predicted for 95% of speakers, which we call the 'speaker variability interval.' As seen before for the group-level pattern, the importance of cues within dialect varies, and furthermore, Jiashan listeners consistently show larger variability than Shanghai listeners for the primary (pitch for Shanghai, contour for Jiashan), the second important (contour for Shanghai, pitch for Jiashan) and the least important cue (both are breathiness), with variance scaling with the mean.

For Shanghai listeners, pitch is always the most important cue for all participants, and higher pitch always leads to more upper register percepts. The findings are also reflected in the individual cue weighting plots (Figure A3 in the Appendix) where pitch always has the highest weighting while the reliance on contour and breathiness varies by individual. For Jiashan listeners, plots of individual cue weights (Figures A3 in the Appendix) further support the larger individual variability for these listeners. Unlike Shanghai listeners, Jiashan listeners show different preferences on the primary cue as well as the use of other cues. Moreover, plots of individual responses also display larger variation for Jiashan listeners (Figures A1 and A2 in the Appendix). It is worth noting that while there seem to be large differences in cue weighting, all participants (except for one participant, 62107, who displayed the opposite categorization) were successful in marking the contrast between the two natural tokens. This suggests that different individuals may have different perceptual strategies, an idea we explore further by inspecting the correlation terms.

Tables 7 and **8** indicate that the models fit some large positive correlations between the cues across individuals. In order to evaluate the significance of the correlations, we extracted the coefficient estimates from the main effects (no interactions) of the fixed effects and the random effects to produce the cue weights from each individual predicted by the model. This yields four estimates per individual, three slopes for three

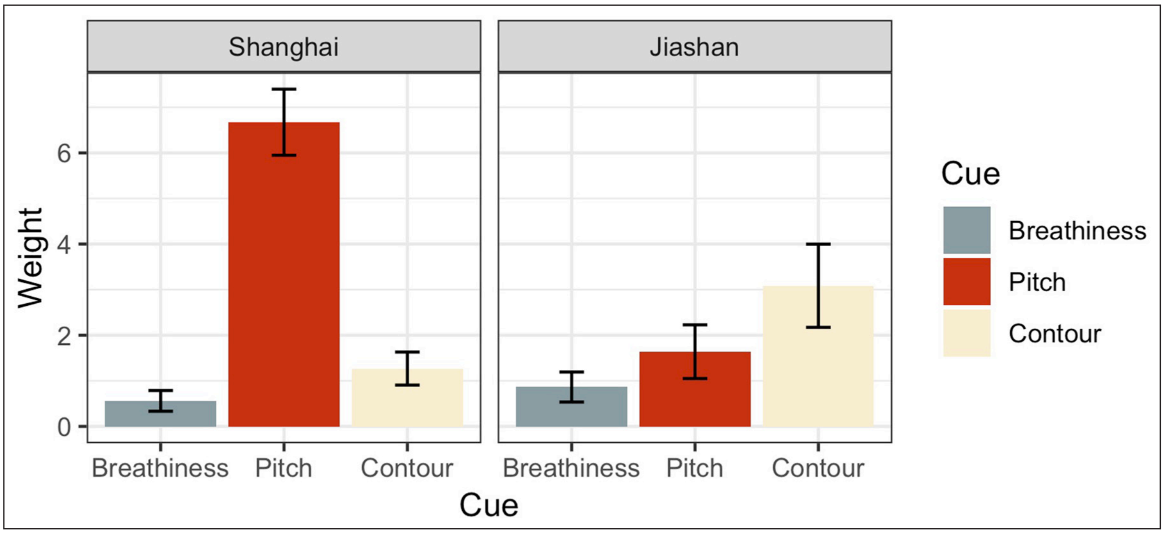

Figure 6: Group-level cue weights for Shanghai (left) and Jiashan (right) listeners with error bars showing 95% speaker variability intervals, capturing individual variation in cue weights. Experiment 1.

cues (breathiness, pitch, contour) and one intercept. We used non-parametric correlation tests using the Spearman method to test all correlations, since non-parametric models are robust to outliers. The results yielded significant correlations among all three cues for the Shanghai listeners (breathiness and pitch: r = 0.87, $p < 0.001$; breathiness and contour: r = 0.87, $p < 0.001$; pitch and contour: r = 0.67, $p < 0.001$) and only between pitch and breathiness for Jiashan listeners (r = 0.94, $p < 0.001$). The relationship between the three cues and the intercepts for each language group is also visually presented in **Figure 7**. **Figure 7** is the result of a Principal Component Analysis (PCA) on the weights calculated for each language separately. PCA is a technique to reduce the dimensionality of the data by transforming possibly correlated variables into a smaller number of uncorrelated variables called 'Principal Components' (Jolliffe, 2002). The first principal component accounts for as much of the variability in the data as possible, and each subsequent component accounts for as much of the remaining variability as possible. **Figure 7** shows that most of the variability in weights for the three dimensions and the intercept can be captured in two dimensions, i.e., the first two principle components (93.68% for Shanghai and 97.12% for Jiashan) and that the two language groups differ in how closely contour patterns with the other two cues and the magnitude of variation in the intercept.

The results show that despite the fact that pitch and breathiness go together (are positively correlated across listeners) for both groups, there are different patterns for the two groups of listeners. Jiashan listeners have a competing primary cue (e.g., participant 62310 has pitch as primary cue and participant 73002 in the same figure has contour; see Figure A2 and A4 in the Appendix), where they may prefer contour or pitch, while for Shanghai listeners, both secondary cues are correlated, and are further correlated with the use of the primary cue (which is always pitch).

In summary, for Jiashan listeners, while contour is the primary cue in aggregate results, this pattern is not representative of all individual listeners. Whether or not a listener relies on contour is independent of their use of pitch and breathiness, while the importance of the other two cues is always positively correlated. For Shanghai listeners, on the other hand, all three cues are correlated, with pitch and breathiness having a stronger correlation.

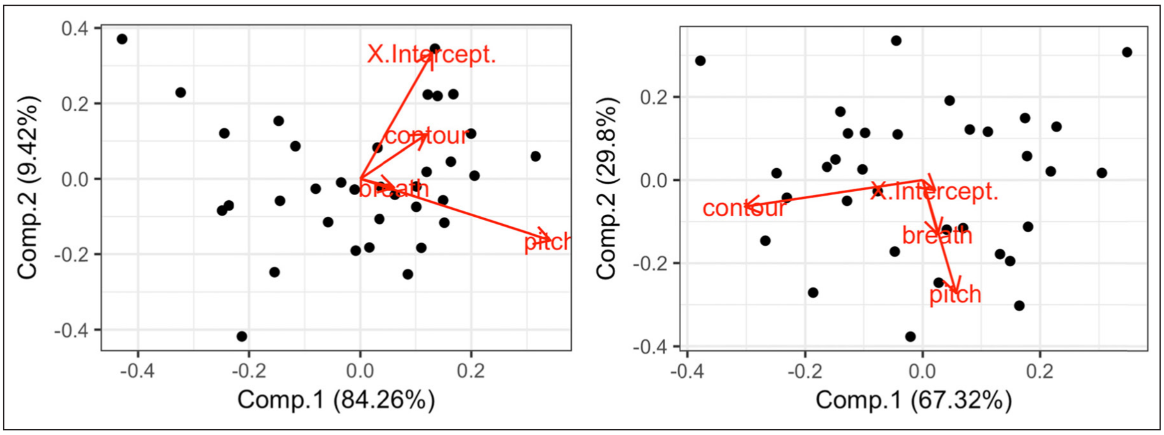

Figure 7: Principal component analysis (PCA) for Shanghai listeners (left) and Jiashan listeners (right) cue weights in Experiment 1. Each dot represents the cue weights of one individual projected onto the first two principal components. The length of the arrows reflects the amount of variation of the term, and the angle between two arrows reflects how much the two terms are correlated (more acute means higher positive correlation; more obtuse means higher negative correlation; the closer to right angle, the more the terms are independent).

3.3.2. Experiment 2

Similar to Experiment 1, this section explores individual variability and the pattern of cue usage using standard deviations reported in the random effects of the mixed-effects models and per-individual coefficient estimates predicted from the main effects (no interactions) of the fixed effects and the random effects. For the purpose of examining the two groups of participants separately, two new models were fit, one for Shanghai participants and one for Jiashan participants, each using the corresponding subset of the data.[9] **Tables 9** and **10** show the summary of random effects for Shanghai and Jiashan participants respectively.

As shown by the standard deviations reported in the two tables, the magnitude of individual variability is similar between the two listener groups. Note also that there is large variation in the effect of talker dialect: Listeners differ more in the effect of listening to different talkers than in the size of any cue.

In terms of correlations among cues, the positive correlation between breathiness and pitch is consistently found in both dialect groups and both experiments (Shanghai: $\mathrm{r} = 0.61, p < 0.001$; Jiashan: $\mathrm{r} = 0.4, p = 0.01$; correlation and significance are from the same Spearman test as Experiment 1), suggesting that listeners' reliance on breathiness

Table 9: Summary of random effects and correlations of Shanghai participants, Experiment 2. Correlations between interaction terms are not shown.

Name	**Variance**	***SD***	**Correlation**			
intercept	0.87	0.93				
Breath	0.13	0.37	0.57			
Contour	0.12	0.35	0.63	0.46		
Pitch	2.13	1.46	0.31	0.48	0.23	
talkerDialect	5.45	2.33	0.13	−0.18	0.49	−0.55
Breath × talkerDialect	0.35	0.59				
Contour × talkerDialect	0.88	0.94				
Pitch × talkerDialect	1.34	1.15				

Table 10: Summary of random effects and correlations of Jiashan participants, Experiment 2. Correlations between interaction terms are not shown.

Name	**Variance**	***SD***	**Correlation**			
intercept	0.62	0.79				
Breath	0.31	0.56	−0.71			
Contour	0.16	0.40	0.52	−0.02		
Pitch	2.19	1.48	0.13	0.40	0.59	
talkerDialect	4.45	2.11	−0.12	−0.06	−0.25	−0.61
Breath × talkerDialect	0.98	0.99				
Contour × talkerDialect	0.99	0.99				
Pitch × talkerDialect	1.56	1.25				

[9] Using the random effects of the original model does not serve this purpose, since there is no random-slope term for listener dialect (there cannot be, as this variable is between-participant). Furthermore, it is impossible (for the class of model fit by blme) to let the random effects structure differ between subsets of the data according to listener dialect in this model.

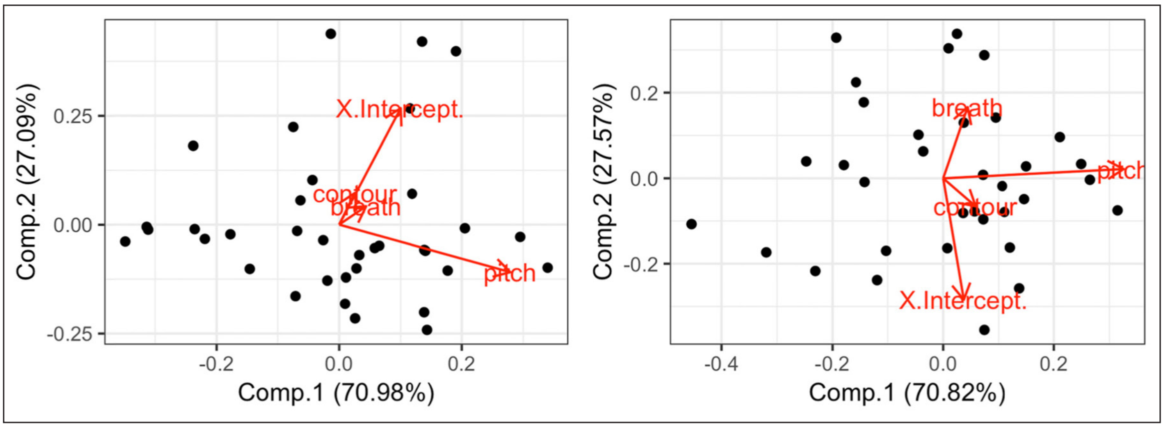

Figure 8: Principle component analysis (PCA) for Shanghai listeners (left) and Jiashan listeners (right), Experiment 2.

is proportional to their reliance on pitch. However, there are some differences between the two experiments. While Shanghai listeners showed a positive correlation between all three cues in Experiment 1, they only show significant correlations between contour and breathiness (r = 0.58, p < 0.001) and between pitch and breathiness in Experiment 2. There is also a trend that all correlations are smaller relative to Experiment 1, with the smallest correlation (i.e., contour and pitch) in Experiment 1 being not significant in Experiment 2. Jiashan listeners, on the other hand, display a positive correlation between pitch and contour (r = 0.69, p < 0.001), which is not found in Experiment 1. Visualization of the Experiment 2 individual variation using PCA, shown in **Figure 8**, further confirms the results of correlations between individuals' cue weights. The small degree of variability in contour weight shown in **Figure 8** probably arises because the checked tones in Jiashan Wu have even shorter duration than those in Shanghai Wu, which makes contour less salient.

In summary, pitch and breathiness are consistently correlated in a positive manner for both groups (although less strongly for Experiment 2). Different from Experiment 1, contour is not correlated with Shanghai listeners' primary cue (i.e., pitch), but instead with the other secondary cue, breathiness. Jiashan listeners, however, display positive correlations between the primary cue and the other two secondary cues.

4. Discussion

This study investigates the role of less important cues (mainly focusing on voice quality) as a way to understand multi-dimensional contrast in Chinese Wu dialects. We compared two genetically-related dialects spoken in close proximity, at the group and individual levels. We considered three research questions, the first on the importance of the secondary cue and its (in)consistency a cross tone pairs, the second on dialectal differences, and the third on individual differences and their implication for the relationship between multiple perceptual cues. We address the first question in 4.1 and 4.2 by discussing the performance of Jiashan listeners (4.1) and Shanghai listeners (4.2), and answer the second question in 4.2 by comparing their different perceptual strategies. We address the third question in 4.3, and address some limitations in interpreting the results in 4.4.

4.1. Jiashan listeners rely on multiple dimensions

The two experiments show that breathiness, while being a secondary cue to the multidimensional tonal register contrast, is nevertheless used in both Shanghai Wu and Jiashan Wu listeners. The difference between the two dialects is that Jiashan listeners' perception is overall more multidimensional with relatively more similar cue weights, while Shanghai listeners' perception is mostly dominated by pitch.

Jiashan listeners as a group have a multidimensional percept of register for the contrasts we studied. This is observed in their relatively large cue weights for the non-primary cues across both experiments, especially when listening to their own dialect. The smaller difference between the cue weights suggests that secondary cues play a bigger role in their perception than for Shanghai listeners.

Jiashan listeners also seem to be more context-dependent in how different cues are used, in this case depending on the particular tone. When comparing the cue weights for falling tones (Experiment 1) and checked tones (Experiment 2), Jiashan listeners use contour as the primary cue for falling tones, but contour is the least important cue for checked tones (likely because of the short syllable duration); meanwhile, the importance of pitch and breathiness is greater in checked tones relative to falling tones. Jiashan listeners also show context dependence in their use of secondary cues when listening to stimuli with different degrees of breathy-modal contrast (Jiashan versus Shanghai talker). Specifically, Jiashan listeners rely more on breathiness when they listened to Jiashan Wu talker—with a larger phonation contrast—than when they listened to Shanghai Wu talker—with a less distinctive contrast in this dimension. It should be noted that these two sets of stimuli vary in more than just the voice quality dimension, as they are different talkers. Nonetheless, what is relevant is that the two listener groups treat the stimuli differently. Moreover, such variation of cue weights under different contexts and speakers/dialects further shows that the secondary cues are indeed learnt from language-specific experience. Shanghai listeners, on the other hand, do not show such variation according to the degree of breathiness in the signal. In fact, their perception strategy appears to be more similar to listeners of languages with no phonation contrasts, a point which we turn to in the next section.

4.2. *Shanghai listeners' perception is dominated by pitch height*

Shanghai listeners' perception of the tonal register contrast differs from Jiashan listeners in that their perception is dominated by pitch, consistent with previous observations (e.g., Zhang & Yan, 2015). In other words, while Shanghai listeners rely on all three cues including breathiness, they exhibit an invariance in their cue weightings for all three sets of stimuli (i.e., Shanghai contour tones, Shanghai checked tones, and Jiashan checked tones): Pitch is always the primary cue and breathiness and contour are always the secondary cues. Furthermore, the secondary cues have quite small cue weights across experiments, especially when listening to the other dialect.

In terms of perception of breathiness, Shanghai listeners may be more like listeners of a language where voice quality is not a cue to lexical contrasts (the third type mentioned in Section 1.2). They seem to rely on voice quality only for the purpose of facilitating pitch location (see Kuang & Liberman, 2015; Kuang et al., 2016 for the influence of voice quality on pitch perception), while Jiashan Wu listeners behave like listeners of a typical language with a phonation contrast. Since there is limited use of breathiness in Shanghai Wu production, it is perhaps not surprising that Shanghai listeners are not very influenced by the breathiness dimension for the Shanghai talker; those stimuli likely do not have very large differences in breathiness, much like stimuli used in past studies (Zhang & Yan, 2015). What is perhaps surprising is that they also are not very influenced by the breathiness dimension for the Jiashan talker in Experiment 2. This provides evidence that these listeners may simply be less sensitive to breathiness, probably using it to facilitate pitch location, indirectly contributing to the tonal register contrast.

Another important piece of evidence supporting the different roles that voice quality plays in Shanghai and Jiashan Wu is the different influence of pitch on breathiness (the breath × pitch interaction terms in **Tables 3** and **4**) in the dialects. At lower pitch, breathiness has a smaller cue weight for Shanghai listeners (compared to higher pitch), but a higher weight for Jiashan listeners. This difference in direction may reflect how voice quality is treated differently in the two dialects: Breathiness, which is always associated with lower pitch, contributes less to the register categorization in Shanghai, as low pitch is already indicative for lower register; on the other hand, high pitch is likely to be perceived lower when produced with breathy voice. In this way breathiness affects the perception of tonal register. The opposite pattern in Jiashan listeners, however, reveals that listeners are more likely to perceive lower register when lower pitch is accompanied by breathiness, suggesting that breathiness is not merely facilitative for pitch location, but directly influences register categorization. Nevertheless, we also acknowledge an alternative explanation, which may require further study: The Jiashan talker has an overall higher pitch range, so breathiness may be more important to identify the lower register, as pitch itself may be ambiguous, while breathiness is less important when the pitch is already high enough to identify a higher register. It is also possible that both these factors contribute to the different pitch-breathiness relationship observed for the two groups of listeners.

We found that the Shanghai listeners were not very influenced by the contour dimension either. However, unlike breathiness which in Experiment 2 was enhanced for the Jiashan talker relative to the Shanghai talker, Experiment 2 did not provide very compelling contour cues in either set of stimuli, limiting our ability to assess use of contour by Shanghai listeners. It is also possible that there is an intrinsic difference between rising tones (heard by the Jiashan listeners of Experiment 1) and falling tones (heard by the Shanghai listeners of Experiment 1), as studies have found that Chinese listeners have smaller Just Noticeable Differences for falling tones than rising tones in terms of both pitch height and pitch contour (Jongman, Qin, Zhang, & Sereno, 2016). Comparison of different tone pairs would be necessary to understand the contribution of contour to the perception of the register contrast in Shanghai Wu.

The above evidence indicates that Shanghai Wu listeners in general rely less on breathiness than Jiashan Wu listeners (and possibly contour as well), regardless of the acoustic signals they heard. Thus, their register contrast seems to be primarily based on pitch height. The decreased cue weight of the voice quality contrast in perception reflects the on-going loss of breathiness in Shanghai speakers' production of the tonal register contrast (Gao, 2016). One possibility is that the difference between the dialects is due to Shanghai listeners' relatively greater contact with Mandarin which does not employ a voice quality contrast (Gao, 2016). Another possibility is that Jiashan listeners' use of more cues in a context-dependent way compared to Shanghai listeners may be because Jiashan Wu has a relatively more complex tone inventory than Shanghai Wu, which may require multiple acoustic dimensions to distinguish between them.

4.3. Individual variability and processing of acoustic cues

We have two major findings on variability between individuals: The variability is highly structured, and there is more variability when the contrast is more multi-dimensional. First, we generally found positive correlations between coefficients of the different cues, and in many cases, these correlations were quite high (more discussion below). Second, in three of our four contrasts (Shanghai rising tones in Experiment 1 and checked tones for all listeners in Experiment 2), the coefficient of the primary cue was much larger than the other two cues in the group data. For these contrasts, the individual data (Figures A3, A5, and A6 in the Appendix) showed that almost all listeners used the same primary cue (with only a few exceptions). In the fourth contrast (Jiashan falling tones in Experiment 1), the group data found that the coefficients for the three cues were more similar, and in the individual data, listeners differed in *which cue* they used as primary. Inspection of the data in Figure A4 in the Appendix shows that 20/33 participants had contour as the primary cue while 11/33 had pitch as the primary cue (the last participant had an ambiguous pattern). This may be due to the falling tone contrast having a more multidimensional nature, with each of the cues providing some, and likely redundant, information. Thus, listeners may come to different solutions to the problem.

We now turn to correlations between individuals' cue coefficients. As has been assumed in previous studies, we think it is correct to assume that group coefficient estimates indicate the degree to which the group uses particular acoustic dimensions. However, we argue that an individual's coefficient estimates are influenced by two factors: the relative importance of a cue with respect to other cues (i.e., the relative magnitude of cues) and the individual's ability to categorize speech stimuli consistently (i.e., overall magnitude, averaged across cues) (see also Kapnoula et al., 2017 for similar arguments). We argue that our analysis shows both of these components and that they have not been considered together before in analyses that report only group-level regression coefficients.

First, when we consider the individual-level coefficients, for both groups of listeners, whenever there is a correlation between two cues, it is always positive. We argue that positive correlations between individuals' coefficients is the default when perception is influenced by more than one cue. This is consistent with previous findings for English showing correlations between regression coefficients for pairs of cues (Clayards, 2018; Hazan & Rosen, 1991; Kim & Clayards, 2019). Furthermore, both Hazan and Rosen (1991) and Clayards (2018) found correlations between the primary cue weights across different contrasts. Together these studies support the idea that some listeners are better able to extract and attend to acoustic phonetic details in a categorization task, regardless of the acoustic dimension. This leads to a more consistent stimulus-response relationship (i.e., sharper categorization functions) and thus larger correlation coefficients.

There are, however, several cases of no significant correlation between cues in the two experiments. Some of these are due to relatively small coefficients of the cues involved, which leads to reduced variability and thus weak correlations (e.g., correlations in Experiment 2 involving contour where contour played only a small role). The more interesting exception is the case of Jiashan listeners' use of pitch and contour in Experiment 1. As discussed above, in this experiment, both cues were important to the contrast and there was very large individual variability in individuals' cue weights. We think the reason for this apparent contradiction comes from the fact that listeners differed in which cue was primary, pitch or contour. Note that previous studies only examine two cues, in which case individuals rarely differ in cue ordering, but in absolute magnitude. Our study, however, focuses on three cues, which provides a venue to examine listeners' cue weights when both *cue ordering* and *weight magnitude* are different across individuals. For the Jiashan listeners in Experiment 1, by definition, those who had contour as a primary cue had smaller cue weights for pitch than contour while those who had pitch as a primary cue had smaller cue coefficients for contour than pitch. Thus, there must be some negative relationship between the cue coefficients for these two cues. We think that this negative relationship in the relative cue coefficients may have been offset by an overall positive relationship between cue coefficients due to individual differences in response consistency (coefficient magnitude) that we observed in all of the other conditions.

In summary, we find that listeners vary in how consistently they respond to cues. Specifically, when listeners agree in cue ordering but only differ in weight magnitude, there is a clear positive correlation; when listeners differ in both cue ordering and weight magnitude, a positive correlation may be masked by their different choice of the primary cue, in which case the correlation pattern may not surface.

4.4. *Limitations*

One may worry whether or not the talkers chosen in the study are representative of the dialect, in other words, whether listeners' responses are specific to the talker, or generalizable for the dialect. We acknowledge that any choice of talker will introduce talker-specific acoustic characteristics that may not be representative for the dialect, and it is hard to verify which specific acoustic dimensions are idiosyncratic. The ideal case would be to use a number of talkers that could jointly be 'representative' of the dialect, although this would be impractical in a single study. In order to eliminate the possibility of talker effects, a follow-up study is needed to explicitly test how listeners react to different talkers. However, even though a single talker cannot stand for the dialect, the acoustics indeed reflect the difference we expect from the two dialects: The Jiashan stimuli have a larger breathy-modal contrast and the Shanghai stimuli have a smaller contrast.

5. Conclusion

This study has examined listeners' perception of a multidimensional tonal register contrast in two Chinese Wu dialects signaled by three cues: pitch height, voice quality, and pitch contour. We found that cue weights are context-specific, i.e., vary by tone. While the cues are present across all cases, some contrasts are more multidimensional than others, as evidenced by cue weights of non-primary cues being bigger for certain tones. In terms of dialectal difference, while both groups of listeners rely on all three cues, Shanghai listeners have smaller cue weights of breathiness than Jiashan listeners. Shanghai listeners rely less on voice quality information even when listening to stimuli with a clear breathy-modal distinction, and their cue weights reflect their dialect-specific experience. Finally, we found structured individual variability: In most cases individuals' cue coefficents were positively correlated, and furthermore, there is more variability across individuals when the contrast is signaled by more than one salient cue, in which case individuals have different options for choosing the primary cue.

Additional File

The additional file for this article can be found as follows:

- **Appendix.** A PDF file containing heat maps for individual responses and individual cue weighting from the two experiments. DOI: https://doi.org/10.5334/labphon.266.s1

Acknowledgements

We thank audiences at the 172nd meeting of the ASA, the 2017 Montreal-Ottawa-Toronto Phonology Workshop, CLA 2017, and LabPhon 2018, as well as Jessamyn Schertz and several anonymous reviewers for their constructive comments. This work was supported by SSHRC grant 435-2016-0747 to Meghan Clayards and SSHRC grants 435-2017-0925 and 430-2014-00018 to Morgan Sonderegger.

Competing Interests

The authors have no competing interests to declare.

References

Abramson, A. S., & Lisker, L. (1985). Relative power of cues: F0 shift versus voice timing. In V. Fromkin (Ed.), *Phonetic Linguistics: Essays in honor of Peter Ladefoged*, 25–33. Orlando: Academic Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3), 255–278. DOI: https://doi.org/10.1016/j.jml.2012.11.001

Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer [Computer program]. Version 6.0.22, retrieved 15 November 2016 from http://www.praat.org/

Brunelle, M. (2009). Tone perception in Northern and Southern Vietnamese. *Journal of Phonetics*, *37*, 79–96. DOI: https://doi.org/10.1016/j.wocn.2008.09.003

Brunelle, M., & Finkeldey, J. (2011). Tone perception in Sgaw Karen. In W.-S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences*, 372–375. Hong Kong: City University of Hong Kong.

Bukmaier, V., Harrington, J., & Kleber, F. (2014). An analysis of post-vocalic/s-ʃ/ neutralization in Augsburg German: Evidence for a gradient sound change. *Frontiers in psychology*, *5*, 828. DOI: https://doi.org/10.3389/fpsyg.2014.00828

Cao, J., & Maddieson, I. (1992). An exploration of phonation types in Wu dialects of Chinese. *Journal of Phonetics*, *20*, 77–92. DOI: https://doi.org/10.1016/S0095-4470(19)30255-4

Chao, Y. R. (1928). Studies in Modern Wu-dialects. Peking: Tsing Hua College Research Institute.

Chao, Y.-R. (1930). A system of tone letters. *Le Maître phonétique*, *45*, 24–27.

Chen, Y., & Gussenhoven, C. (2015). Shanghai Chinese. *Journal of the International Phonetic Association*, *45*(3), 321–337. DOI: https://doi.org/10.1017/S0025100315000043

Chen, Z. (2015). Breathy voice and low tone. *Journal of Chinese Linguistics*, *43*(1), 90–117. DOI: https://doi.org/10.1353/jcl.2015.0004

Chen, Z. M. (2010). An Acoustic Study of Voiceless Onset Followed by Breathiness of Wu Dialects: Based on the Shanghai Dialect. *Studies in Language and Linguistics*, *30*(3), 20–34.

Chung, Y., Rabe-Hesketh, S., Dorie, V., Gelman, A., & Liu, J. (2013). A nondegenerate penalized likelihood estimator for variance parameters in multilevel models. *Psychometrika*, *78*(4), 685–709. http://gllamm.org/. DOI: https://doi.org/10.1007/s11336-013-9328-2

Clayards, M. (2018). Individual differences in speech perception cue weights are correlated within and across contrasts. *Journal of the Acoustical Society of America Express Letters*, *144*(3). DOI: https://doi.org/10.1121/1.5052025

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. DOI: https://doi.org/10.1016/j.wocn.2017.09.009

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. DOI: https://doi.org/10.1121/1.2945161

Gao, J. (2016). Sociolinguistic motivations in sound change: On-going loss of low tone breathy voice in Shanghai Chinese. *Papers in Historical Phonology*, *1*, 166–186. DOI: https://doi.org/10.2218/pihph.1.2016.1698

Gao, J., Hallé, P., & Draxler, C. (2019). Breathy voice and low-register: A case of trading relation in Shanghai Chinese tone perception? *Language and Speech*. DOI: https://doi.org/10.1177/0023830919873080

Gao, J., Hallé, P., Honda, K., Maeda, S., & Toda, M. (2011). Shanghai slack voice: Acoustic and EPGG data. In W.-S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress on Phonetic Sciences*, 719–722. Hong Kong: City University of Hong Kong.

Garellek, M., & Keating, P. (2011). The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *Journal of the International Phonetic Association*, *41*(2), 185–205. DOI: https://doi.org/10.1017/S0025100311000193

Garellek, M., Keating, P., Esposito, C. M., & Kreiman, J. (2013). Voice quality and tone identification in White Hmong. *The Journal of the Acoustical Society of America*, *133*(2), 1078–1089. DOI: https://doi.org/10.1121/1.4773259

Gordon, P. C., Eberhardt, J. L., & Rueckl, J. G. (1993). Attentional modulation of the phonetic significance of acoustic cues. *Cognitive Psychology*, *25*(1), 1–42. DOI: https://doi.org/10.1006/cogp.1993.1001

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

Hollien, H. (1974). On Vocal Registers. *Journal of Phonetics*, *2*, 125–143. DOI: https://doi.org/10.1016/S0095-4470(19)31188-X

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

Honorof, D. N., & Whalen, D. H. (2005). Perception of pitch location within a speaker's F0 range. *The Journal of the Acoustical Society of America*, *117*(4), 2193–2200. DOI: https://doi.org/10.1121/1.1841751

Jiang, B., & Kuang, J. (2016). Consonant effects on tonal registers in Jiashan Wu. *Proceedings of the Linguistic Society of America*, *1*(30), 1–13. DOI: https://doi.org/10.3765/plsa.v1i0.3730

Jolliffe, I. T. (2002). *Principal Component Analysis*. Second edition. New York: Springer-Verlag.

Jongman, A., Qin, Z., Zhang, J., & Sereno, J. (2016). Just noticeable differences for pitch height and pitch contour for Chinese and American listeners. *The Journal of the Acoustical Society of America*, *140*(4), 3225–3225. DOI: https://doi.org/10.1121/1.4970184

Kapnoula, E. C., Winn, M. B., Kong, E. J., Edwards, J., & McMurray, B. (2017). Evaluating the sources and functions of gradiency in phoneme categorization: An individual differences approach. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(9), 1594–1611. DOI: https://doi.org/10.1037/xhp0000410

Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., & Banno, H. (2008). Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, 3933–3936. IEEE. DOI: https://doi.org/10.1109/ICASSP.2008.4518514

Kim, D., & Clayards, M. (2019). Individual differences in the link between perception and production and the mechanisms of phonetic imitation. *Language, Cognition and Neuroscience*, 1–18. DOI: https://doi.org/10.1080/23273798.2019.1582787

Kingston, J. (2011). Tonogenesis. In M. van Oostendorp (Ed.), *The Blackwell Companion to Phonology*, 1–30. New York: John Wiley & Sons. DOI: https://doi.org/10.1002/9781444335262.wbctp0097

Kirby, J. P. (2014). Incipient tonogenesis in Phnom Penh Khmer: Acoustic and perceptual studies. *Journal of Phonetics*, *43*, 69–85. DOI: https://doi.org/10.1016/j.wocn.2014.02.001

Kong, E. J., & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics*, *59*, 40–57. DOI: https://doi.org/10.1016/j.wocn.2016.08.006

Kong, E. J., & Lee, H. (2018). Attentional modulation and individual differences in explaining the changing role of fundamental frequency in Korean laryngeal stop perception. *Language and Speech*, *61*(3), 384–408. DOI: https://doi.org/10.1177/0023830917729840

Kuang, J. (2011). *Production and perception of phonation contrasts in Yi*. MA thesis, University of California Los Angeles.

Kuang, J. (2013). The tonal space of contrastive five level tones. *Phonetica*, *70*(1–2), 1–23. DOI: https://doi.org/10.1159/000353853

Kuang, J., Guo, Y., & Liberman, M. (2016). Voice quality as a pitch-range indicator. *Proceedings of Speech Prosody 2016*. Boston. DOI: https://doi.org/10.21437/SpeechProsody.2016-218

Kuang, J., & Liberman, M. (2015). The effect of spectral slope on pitch perception. In *Sixteenth Annual Conference of the International Speech Communication Association*, 354–358.

Lenth, R., Love, J., & Hervé, M. (2017). Package 'emmeans'. *Statistician*, *34*(4), 216–221.

Lisker, L. (1978). In qualified defense of VOT. *Language and Speech*, *21*(4), 375–383. DOI: https://doi.org/10.1177/002383097802100413

Mayo, C., & Turk, A. (2004). Adult–child differences in acoustic cue weighting are influenced by segmental context: Children are not always perceptually biased toward transitions. *The Journal of the Acoustical Society of America*, *115*(6), 3184–3194. DOI: https://doi.org/10.1121/1.1738838

Mayo, C., & Turk, A. (2005). The influence of spectral distinctiveness on acoustic cue weighting in children's and adults' speech perception. *The Journal of the Acoustical Society of America*, *118*(3), 1730–1741. DOI: https://doi.org/10.1121/1.1979451

McAuliffe, M., & Babel, M. (2016). Stimulus-directed attention attenuates lexically-guided perceptual learning. *The Journal of the Acoustical Society of America*, *140*(3), 1727–1738. DOI: https://doi.org/10.1121/1.4962529

Nicenboim, B., & Vasishth, S. (2016). Statistical methods for linguistic research: Foundational Ideas—Part II. *Language and Linguistics Compass*, *10*(11), 591–613. DOI: https://doi.org/10.1111/lnc3.12207

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review*, *85*(3), 172. DOI: https://doi.org/10.1037/0033-295X.85.3.172

Qian, N. R. (1992). Studies on Contemporary Wu. Shanghai: Shanghai Education Press.

Raphael, L. J. (2005). Acoustic cues to the perception of segmental phonemes. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception*, 182–206. Malden, MA: Blackwell. DOI: https://doi.org/10.1002/9780470757024.ch8

Ren, N. (1992). *An acoustic study of Shanghai stops*. Doctoral dissertation, University of Connecticut.

Sadakata, M., & Sekiyama, K. (2011). Enhanced perception of various linguistic features by musicians: A cross-linguistic study. *Acta psychologica*, *138*(1), 1–10. DOI: https://doi.org/10.1016/j.actpsy.2011.03.007

Schertz, J., & Clare, E. J. (2019). Phonetic cue weighting in perception and production. *Wiley Interdisciplinary Reviews: Cognitive Science*, e1521. DOI: https://doi.org/10.1002/wcs.1521

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. DOI: https://doi.org/10.1121/1.4736711

Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics*, *71*, 147–161. DOI: https://doi.org/10.1016/j.wocn.2018.07.008

Wayland, R., & Jongman, A. (2003). Acoustic correlates of breathy and clear vowels: The case of Khmer. *Journal of Phonetics*, *31*(2), 181–201. DOI: https://doi.org/10.1016/S0095-4470(02)00086-4

Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. (1993). F0 gives voicing information even with unambiguous voice onset times. *The Journal of the Acoustical Society of America*, *93*(4), 2152–2159. DOI: https://doi.org/10.1121/1.406678

Xu, B., & Tang, Z. (1988). *A description of the urban Shanghai dialect*. Shanghai: Shanghai Educational Publishing House.

Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9781139164559

Yip, M. J. (1980). *The tonal phonology of Chinese*. Doctoral dissertation, Massachusetts Institute of Technology.

Yonezawa, T., Suzuki, N., Mase, K., & Kogure, K. (2005, May). HandySinger: Expressive singing voice morphing using personified hand-puppet interface. In *Proceedings of the 2005 Conference on New Interfaces for Musical Expression*, 121–126. Singapore: National University of Singapore.

Yu, G. (1988). Syllabary of homophones in Jiaxing dialect. *Dialect*, *3*, 195–208.

Yu, A. C., & Zellou, G. (2019). Individual Differences in Language Processing: Phonology. *Annual Review of Linguistics*, *5*, 131–150. DOI: https://doi.org/10.1146/annurev-linguistics-011516-033815

Yu, K. M., & Lam, H. W. (2011). The role of creaky voice in Cantonese tonal perception. In W.-S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences*, 2240–2243. Hong Kong: City University of Hong Kong.

Zhang, J., & Yan, H. (2015). Contextually dependent cue weighting for a laryngeal contrast in Shanghai Wu. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow.

How to cite this article: Jiang, B., Clayards, M., & Sonderegger, M. 2020 Individual and dialect differences in perceiving multiple cues: A tonal register contrast in two Chinese Wu dialects. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 11(1): 11, pp. 1–30. DOI: https://doi.org/10.5334/labphon.266

Submitted: 13 February 2020 **Accepted:** 07 July 2020 **Published:** 21 August 2020

]u[*Laboratory Phonology: Journal of the Association for Laboratory Phonology* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS