Humanities and Social Sciences Communications

Article in Press

https://doi.org/10.1057/s41599-026-07071-9

Individual differences in phonation types and their interaction with pitch range: Evidence from the five level tones in Hmu

Received: 7 October 2025

Accepted: 12 March 2026

Cite this article as: Liu, W., Hou, N., Tang, H. Individual differences in phonation types and their interaction with pitch range: Evidence from the five level tones in Hmu. *Humanit Soc Sci Commun* (2026). https://doi.org/10.1057/s41599-026-07071-9

Wen Liu, Nianhan Hou & Hao Tang

Individual differences in phonation types and their interaction with pitch range: Evidence from the five level tones in Hmu

Abstract: Cross-linguistic studies on tone and phonation have revealed the role of laryngeal phonatory settings in tonal contrasts. However, systematic research on how individuals within the same speech community achieve multidimensional tonal distinction remains lacking. Hmu, an Eastern Hmongic language, is typologically notable for its five level tones, offering ideal material for examining the interaction between pitch and phonation. Based on acoustic and EGG data collected from 30 speakers, this study investigates the differentiation strategies of the five level tones at both the group and individual levels. The results reveal that T11 differs significantly from the other four level tones across acoustic and EGG parameters, characterized by a larger spectral tilt and a higher noise level, aligning with the properties of breathy voice. In contrast, the other four level tones generally exhibit a smaller spectral tilt and lower noise, consistent with the characteristics of modal voice. Among them, T55, due to its high F0, may be further identified as a high-pitched voice. Individually, native speakers show variation in how they utilize phonations to encode linguistic contrast in T11, with three primary subtypes observed: breathy, harsh, and near-modal voice. Significantly, the non-modal phonation associated with T11 does not extend across the entire vowel but is primarily concentrated in the first third. We also found that an individual's pitch range may be one factor influencing the number of acoustic cues they use when distinguishing tones. Speakers with a narrower pitch range usually employ non-modal phonations. This study provides empirical evidence that tonal contrast is multidimensional and offers a referential analysis method for future investigations into individual variation in phonation type.

Keywords: Hmu; five level tones; phonation types; individual differences

1 Introduction

Hmu (Xinzhai variety; henceforth Hmu) is spoken by approximately 1,000 people residing in Xinzhai Village, Kaili City, Guizhou Province, China. It is an Eastern Hmongic language of the Hmong-Mien language family (Liu et al., 2020). Hmu has a complex tonal system that includes changes in pitch as well as non-modal phonation. One typologically striking feature of Hmu is that it has five contrastive level tones: T11 (low level), T22 (mid-low level), T33 (mid level), T44 (mid-high level), and T55 (high level). To our knowledge, languages with five level tones are extremely rare in the known languages (Chao, 1948; Maddieson, 1978; Kuang, 2013), making Hmu an important case for understanding how dense tonal contrasts are maintained in production and perception.

The rarity of five-level-tone system raises a fundamental question: how can such a crowded tonal inventory remain perceptually distinct and phonologically stable? This question can be examined through the framework of Dispersion Theory (Liljencrants and Lindblom, 1972; Lindblom, 1986, 1990; Flemming, 1995). According to Dispersion Theory, there are two strategies to maximize phonemic contrast among languages. One way is to expand the overall acoustic space of the inventories. Cross-linguistic studies showed that a positive correlation exists between the number of vowels (Lindblom, 1986; Schwartz et al., 1997a, 1997b; Becker-Kristal, 2010) or tones (Maddieson, 1978; Kuang, 2013) and the size of their acoustic space. The second way is to add contrast

dimensions (Lindblom and Maddieson, 1988). In tonal contrasts, phonation has been widely used as one such additional dimension that can enhance the distinctiveness of tones, especially when pitch-based contrasts are crowded.

Hmongic languages provide a fertile testing ground for examining the interaction between pitch and phonation in tonal contrast. In addition to pitch differences, tones are usually associated with non-modal phonations. Of these, breathy voice is the most common type and occurs with the mid-falling tone in Green Mong (Huffman, 1987; Andruski and Ratliff, 2000), the high-falling tone in White Hmong (Esposito, 2012; Garellek et al., 2013), the mid-level tone in Black Miao (Kuang, 2013), and the low-level tone in Hmu (Liu et al., 2020, 2024). Besides, the low falling tone in Green Mong (Huffman, 1987; Andruski and Ratliff, 2000) and White Hmong (Esposito, 2012; Garellek et al., 2013) is accompanied by creaky voice; Black Miao uses tense voice in the high level tone and vocal fry in the low level tone (Kuang, 2013).

Regarding Hmu, based on preliminary analysis (Liu, 2020; Liu et al., 2020, 2024), the five level tones can be well-distinguished by native speakers, and T11 is accompanied by non-modal phonation; moreover, the perceptual experiment further confirmed that native speakers rely on the non-modal phonation when identifying T11, indicating that phonation contributes meaningfully to tonal contrast in this highly crowded system. Despite these insights, limited by speakers and acoustic parameters, the resulting findings were tentative, and the phonation types of the five level tones and the intra-syllabic positioning of T11's phonation were neither rigorously defined nor thoroughly analyzed using multidimensional acoustic and EGG measures. Additionally, inter-speaker phonation variation in T11 has not been addressed.

This gap is particularly significant from the perspective of Dispersion Theory. In a crowded tonal system such as Hmu, it remains unclear whether phonation cues are employed uniformly across speakers or whether individuals adopt different strategies depending on physiological constraints, such as vocal pitch range. Against this background, this study mainly addresses three questions: (i) determining the phonation categories of the five level tones and their fine-grained temporal phonation variation within syllables; (ii) examining individual differences among speakers when producing T11; (iii) exploring the interaction between phonation and pitch range for distinguishing the five level tones. By answering these questions, this study contributes to a more nuanced understanding of how multidimensional tonal contrasts are realized in languages with crowded level tone and highlights the importance of individual variation in phonological implementation. The following section (1.1–1.3) will review the relevant research background around these three issues.

1.1 Phonation types across languages

Phonation is the production of sound by the vocal folds, which is often used in a narrow sense to refer only to the production of voicing, i.e., the rate (perceived as pitch) and the manner (perceived as phonation types) of vocal fold vibration. Phonation has attracted much attention over the last two decades, and the primary objective of studying it is to explore the relationship between the sounds produced by the vocal folds and their linguistic meanings.

Typically, phonation varies along a glottal openness continuum (i.e., degrees of glottal width defined by apertural variance between the arytenoid cartilages), and glottal openness refers to the proportion of the glottal open phase to a single glottal vibration cycle. The most common phonations are modal voice, creaky voice, and

breathy voice (Ladefoged, 1971; Gordon and Ladefoged, 2001; Brunelle and Kirby, 2016; Garellek, 2019, 2022; Esposito and Khan, 2020; Liu, 2021; Keating et al., 2023). Modal voice serves as the neutral mode of phonation: the vibration of vocal folds is periodic without any audible frication, and the overall laryngeal tension is moderate (Laver, 1980). Modal voice is the basis for defining other phonation types, and the non-modal voice is an inclusive term for any phonation type that deviates from the modal voice (Liu, 2021). Tense voice is sometimes considered intermediate to creaky and modal voice, while lax voice is between breathy and modal voice (Keating et al., 2011; Kuang and Keating, 2014). Additionally, other terms, such as "strident" and "harsh" are utilized to describe non-modal phonations involving supraglottalic mechanisms or epilaryngeal constriction in addition to changes in vocal fold vibration (Laver, 1980; Ladefoged and Maddieson, 1996; Traill, 1986; Gerratt and Kreiman, 2001; Esling and Harris, 2005; Edmondson and Esling, 2006; Miller, 2007; Moisik and Esling, 2011, 2014; Moisik, 2013; Moisik et al., 2014, 2021).

Phonation can usually be divided into modal and non-modal categories. To date, a fully parameterized definition of phonations has not been achieved. Existing criteria are inherently relative, relying on the modal phonation as a reference (Laver, 1980). Relative to modal voice, the other phonations can be collectively categorized as non-modal phonations. Once the classification of phonations is established, the next challenge is determining how to categorize or characterize these phonation categories. Two basic dimensions of voice quality—spreading/constriction and noise—have specific acoustic attributes in the psychoacoustic model (Kreiman et al., 2014); accordingly, the spectral tilt and noise measures are the most commonly used (Garellek, 2019).

1.2 Acoustic and EGG measures of phonation

Acoustically, many parameters have been identified in past literature to quantify phonation (Gordon and Ladefoged, 2001; Garellek, 2019, 2022; Esposito and Khan, 2020; Liu, 2021; Chai and Garellek, 2022; Keating et al., 2023).

The most common spectral tilt measure is the amplitude difference between the first and the second harmonics (H1-H2), which reflects the open phase duration during each glottal cycle (Holmberg et al., 1995). In addition, other spectral tilt measures include the amplitude difference between the second and fourth harmonics (H2-H4: Kreiman et al., 2007), and the amplitude between the first harmonic and the harmonics exciting higher formants (H1–A1, H1–A2, H1–A3). H2–H4 is relevant to listeners' identification of the speaker's sex (Bishop and Keating, 2012). H1–A1, H1–A2, and H1–A3 are reported to reflect changes in the strength of higher frequencies, and H1–A3 is often considered to indicate the abruptness of vocal fold closure (Stevens, 1977). Compared to modal voice, higher values of these spectral tilt measures indicate a breathier phonation, while lower values indicate a creakier one. Furthermore, measures of periodicity and harmonics-to-noise ratio have also been used to quantify phonation differences. A common measure is cepstral peak prominence (CPP: Hillenbrand et al., 1994), and it is an indicator of contrastive breathy phonation. Another spectral noise measure is harmonics-to-noise ratio (HNR: de Krom, 1993), and it is also useful for distinguishing breathy/non-breathy

and creaky/non-creaky voice. Usually, CPP and HNR are higher for modal voice and lower in both breathier and creakier phonations, due to weaker harmonics and stronger noise components in the speech signal.

Electroglottography (EGG) is also a useful tool in measuring vocal fold contact area during phonation (Fabre, 1957; Frokjaer-Jensen and Thorvaldsen, 1968; Fourcin and Abberton, 1971). The most common EGG measure is the contact quotient (CQ), defined as the ratio of the contact phase duration to the pitch period duration (Rothenberg and Mahshie, 1988; Baken and Orlikoff, 2000). CQ is inversely correlated with H1-H2, indicating that both reflect glottal aperture (DiCanio, 2009; Esposito, 2012). Breathy voice tends to have much lower CQ and greater H1-H2 values than its modal counterpart, while creaky voice produced with greater vocal fold contact has greater CQ and lower H1-H2 values than its modal counterpart. Another common measure, Derivative-EGG Closure Peak Amplitude (DECPA), also called Peak Increase in Contact (PIC), is defined as the amplitude of the positive peak on the Derivative-EGG (i.e., the highest rate of increase of vocal fold contact), and has been used to measure the speed of glottal closure. Generally, phonations produced with faster glottal closure have greater DECPA or PIC values than phonations produced with slower glottal closure.

Additionally, non-modal phonations also affect formants. Compared to their modal counterparts, breathy or lax voice is usually accompanied by a lower F1, such as in Dinka (Denning, 1989), Xhosa (Jessen and Roux, 2002), Southern Yi (Kuang and Cui, 2018), Chrau (Tạ et al., 2022). In contrast, creaky or tense voice often has a higher F1, such as in Hani (Maddieson and Ladefoged, 1985), Mpi (Blankenship, 2002), and Yi (Kong, 2001). Similarly, vowels with harsh voice also tend to have a higher F1, as in Fuzhou Min Chinese (Chen and Havenhill, 2025).

1.3 Phonation variation in speech production

Similar to many phonological features, phonation also varies within and across languages. Languages with phonation contrasts can exhibit distinctive timing patterns, with non-modal phonations occurring on consonants and/or vowels (see Gordon and Ladefoged, 2001; Esposito and Khan, 2020; Garellek, 2022, for reviews), with the domain shifting dynamically within the syllable (Silverman et al., 1995; Blankenship, 1997; Kong, 2001; Liu, 2021). Compared to consonants, non-modal phonations mainly occur on vowels (Keating et al., 2023). Besides, non-modal phonation is typically localized to specific portions of the vowels, occurring at the beginning (Jalapa Mazatec: Silverman, 1997; Blankenship, 2002; Keating et al., 2010; Garellek and Keating, 2011; Garellek et al., 2021; Chanthaburi Khmer: Wayland and Jongman, 2003; Hmu: Liu et al., 2024), the middle (Keating et al., 2023), and the end (SAV Zapotec: Esposito and Khan, 2020).

Moreover, even within the same language, phonations also vary across individuals (Kreiman et al., 2007; Bishop and Keating, 2012; Lee et al., 2019). For example, creaky or laryngealized vowels manifest as audible creak, subtle laryngealization, or without any audible creakiness across speakers in Coatzospan Mixtec (Gerfen and Baker, 2005). In Bai, speakers may use harsh or high-pitched voice for the high falling tone (Li and Wang, 2016), and breathy or tense voice for the low falling tone (Liu et al., 2019). In Hmu, breathy, harsh, and nearly modal voice have been observed in the low level tone (Liu et al., 2020). Fuzhounese speakers produce the low-register tones (tone values: /21/, /241/, /24/) with creaky, ventricular, whispery, and harsh (Chen and Havenhill, 2025).

2. Methods

2.1. Materials

Based on our fieldwork, Hmu minimal tone sets with eight tones (108 monosyllables) were used in the production experiment (Table S1). The initial consonants of the sample words include unaspirated stop, unaspirated affricate, and lateral (/p, t, k, q, ts, tɕ, l/), and the finals include monophthongs (/i, ɛ, a, ə, o, u/) and diphthongs (/ia, iu, iə/). A total of 6480 tokens were obtained.

2.2. Speakers

Thirty native Hmu speakers were recruited from Xinzhai village, including 20 males (M01, M02, M03, M04, M05, M06, M07, M08, M09, M10, M11, M12, M13, M14, M15, M16, M17, M18, M19, M20; age range: 26 – 58 years) and 10 females (F01, F02, F03, F04, F05, F06, F07, F08, F09, F10; age range: 18 – 43 years). All speakers were bilingual in Hmu and Southwest Mandarin. However, Hmu remained the primary language of oral communication at family and community events. All participants provided informed consent in accordance with a protocol approved by the Survey and Behavioural Research Ethics Committee of Shandong University.

2.3. Recordings

Simultaneous audio and electroglottographic (EGG) signals were recorded using a Sony ECM-44B clip-on microphone as the first channel and Glottal Enterprises Electroglottograph (model EG2) as the second channel. Given that the recording was conducted in a field environment, to control environmental variables, all recordings were completed in the same quiet room in Xinzhai village to ensure consistency with the recording environment. The microphone with a windscreen was placed on the left side of the chest, approximately 15–21 cm from the speaker's mouth. When placing EGG electrodes, first cover the gold surfaces of the electrodes with a very thin coating of water-soluble electrode gel, then fit the electrodes tightly against the neck to fix them to both sides of the thyroid cartilage. A trial recording with the continuous vowel /a/ was conducted before each participant's recording, and the signal quality was assessed on-site until the "Electrode Placement/Laryngeal Movement" indicator on the hardware's front panel was centered on the green LEDs and relatively stable. The recording was made using Adobe Audition 2.0 at a sampling rate of 44.1 kHz and 16-bit amplitude resolution.

Given that Hmu lacks a written system, the experiment stimuli were presented using Chinese characters. All participants were bilingual, enabling them to understand word meanings from Chinese characters and then read the target words in Hmu. The stimuli were further refined based on the authors' previous linguistic fieldwork, which used a 3000-word vocabulary list, ensuring that each Chinese character corresponded to a commonly used, unambiguous Hmu word. The Hmu speakers were asked to read the monosyllabic words on the paper naturally and comfortably. Each token was repeated twice, with a 2-second interval between repetitions. So, there were 216 items for each speaker, and 6480 tokens (108 sample words × 2 repetitions × 30 native speakers) were obtained. The words were spoken in isolation to avoid tone coarticulation in continuous speech. Elicitation was

carried out in Hmu by the linguistic consultant to ensure that there was no interference from Southwestern Mandarin. Meanwhile, during the recording, the linguistic consultant monitored their utterances to check that tones were produced accurately and that repetitions were fluent.

2.4. Measures and Data analysis

Based on the overview in the Section 1.2, the most widely used parameters for tone and phonation contrast were chosen, including acoustic measures (F0, duration, H1*-H2*, H1*-A1*, H1*-A2*, H1*-A3*, CPP, HNR05, HNR15, HNR25, HNR35) and EGG measures (CQ, DECPA). These parameters have been shown to exhibit good sensitivity and stability in distinguishing modal and non-modal phonations across languages. Note that the four spectral tilt measures were corrected with vowel formants (Iseli et al., 2007). For convenience, H1*-An* was used to refer to H1*-A1*, H1*-A2*, and H1*-A3*; similarly, HNRs was used to refer to HNR05, HNR15, HNR25, and HNR35. The eleven acoustic measures were extracted automatically from audio signals using VoiceSauce (Shue et al., 2011). It should be noted that although loudness/subglottic pressure may affect spectral parameters such as H1-An, this study asked subjects to maintain a natural and comfortable phonation state, and did not include the extraction or comparison of loudness/subglottic pressure; therefore, the sound pressure level of the recording was not calibrated.

All EGG signals undergo a standardized preprocessing procedure that uses FFT high-pass filtering to eliminate low-frequency baseline drift caused by laryngeal movement, while filtering out low-frequency motion artifacts unrelated to changes in vocal fold contact, thereby reducing signal noise. Then, the EGG signals corresponding to the audio tokens were processed by EggWorks (Tehrani, 2009). In this procedure, CQ was measured using the popular "Hybrid" method (Howard, 1995), in which the onset of the contact phase is defined by the positive peak in the derivative EGG signal and the offset by an amplitude threshold in the raw EGG signal, with the threshold set to 25% of maximal amplitude. This method fully accounts for the difficulty of determining the glottal opening instant and has been widely used in the literature (e.g., Esposito, 2012; Garellek, 2022). DECPA is the positive peak value from the derivative EGG signal (dEGG). The positive peak in the dEGG signal corresponds to the maximum speed of vocal fold closure and is a measure of the speed of the vocal folds at the moment of closure.

For each token, we first manually annotated the vowel (including monophthongs and diphthongs). The annotation began at the second cycle of vocal fold vibration, and the endpoint was determined based on F0 (regular vocal fold vibration), formants (stable structure), and intensity (no significant attenuation), to ensure that the analyzed segment was stabilized. The labeled vowel was divided into nine equal-duration parts for plotting the tone curve; the mean values for each third of each labeled segment interval were also obtained. Measurements were made at all parts by averaging the value (for a given measure) of this part and then averaged across the labeled segment. In this study, to determine the localization of non-modal phonation, we divide the vowel into three segments for statistical testing. In the data preprocessing, outliers are removed using the interquartile range (IQR) method.

3. Results

3.1. Acoustic characteristics of the five level tones

Hmu has eight phonemic tones. Fig. S1 demonstrates the average F0 curves of male and female speakers. Overall, the tones are well-distinguished in tonal space and can be grouped by tone contour: five level tones (T55, T44, T33, T22, T11), two rising tones (T24, T23), and one falling tone (T51). The pitch range of males (134.04 Hz) is slightly wider than that of females (105.75 Hz) ($p > 0.05$). In terms of duration, rising tones are longest, followed by level tones, whereas the falling tone is shortest.

Fig. S1 is to be inserted here

The following section focuses on the five level tones. Fig. 1 shows the average F0 curves across 30 speakers. While the five level tones show a consistent general trend, considerable inter-speaker variation is observed. Certain speakers (e.g., F04, F06, M02, M09, M19) exhibit a wider pitch range and more distinct F0 differences across the five level tones. Others show limited distinction between T11 and T22. This is particularly evident in individuals with a narrow pitch range, where F0 curves of T11 and T22 overlap (e.g., F05, F09, F10). In other cases, F0 curves of T11 and T22 exhibit slight separation at the onset but converge in the latter half (e.g., M08, M13).

Fig. 1 is to be inserted here

Fig. 2 presents the distribution of acoustic and EGG parameters by gender. As pitch increases, all speakers exhibit a decreasing trend in H1*-An*. T11 has a slightly higher H1*-H2* among males compared with the other level tones. In contrast, the noise parameters exhibit a distinct pattern for both males and females: CPP and HNRs of T11 are significantly lower than those of the other level tones. Regarding EGG measures, T11 is characterized by lower CQ and DECPA in males; females also show lower DECPA for T11.

Fig. 2 is to be inserted here

To further examine phonation characteristics of the five level tones, we constructed linear mixed-effects models (LMMs) using the lmerTest package in R4.4.2 (Kuznetsova et al., 2017). All measures were selected as the dependent variables. Tone category and gender were considered fixed effects, while speaker and CV were considered random effects. The results showed that the main effects were significant for all parameters (ps < 0.001). Meanwhile, the interaction effects between tone and gender were also significant for all parameters (ps < 0.001), suggesting the differences in tone categories are modulated by gender. Given that the interaction effects of all parameters were significant, we further performed a simple main effects analysis. The results showed that the main effects of tone were significant at both gender levels. Subsequently, we performed post hoc multiple comparisons with the Bonferroni correction for each tone category within each gender (Table S2).

As shown in Table S2, for male speakers, T11 differs significantly from the other four level tones. Among female speakers, similar significant differences are also observed for T11 except CQ. Meanwhile, the differences

between T11 and the other four level tones are larger than the differences among the remaining four level tones themselves. Combined with acoustic and EGG data, relative to the other four level tones, T11 is distinguished by larger H1*-H2* and H1*-An*, lower CPP, HNRs, CQ, and DECPA, indicating that T11 has a steeper spectral tilt, higher noise level, increased glottal leakage, and slower vocal fold closure, aligning with the phonatory properties of breathy voice. In contrast, T22, T33, T44, and T55 demonstrate smaller H1*-H2* and H1*-An*, higher CPP, HNRs, CQ, and DECPA, generally conforming to the characteristics of modal voice. In addition to its significant distinction from T11, T55 also differs significantly from T22, T33, and T44 in H1*-An*. The values of H1*-An* for T55 are lower than those of the other level tones, suggesting that higher pitch is associated with a faster vocal fold closure, exhibiting the phonation characteristics of high-pitched voice.

Table 1 Post-hoc pairwise comparisons across vowel intervals (onset: 01; mid: 02; offset: 03) (estimated values; and Bonferroni-adjusted significance indicated by asterisks: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$).

Variable	Gender	01 vs. 02	01 vs. 03	02 vs. 03
H1*-H2*	M/F	0.93***	0.96***	0.02
H1*- A1*	M	1.93***	1.05***	-0.88***
	F	0.70	-0.57	-1.28***
H1*- A2*	M	1.99***	0.36	-1.41***
	F	0.32	-1.27**	-1.59***
H1*- A3*	M	2.48***	1.13***	-1.35***
	F	0.43	-0.36	-0.79
CPP	M	-3.89***	-3.38***	0.51***
	F	-4.10***	-2.75***	1.34***
HNR05	M	-11.23***	-15.00***	-3.76***
	F	-12.08***	-10.60***	1.48**
HNR15	M	-9.03***	-13.40***	-4.37***
	F	-8.54***	-8.07***	0.48
HNR25	M	-9.23***	-13.90***	-4.68***
	F	-8.44***	-8.92***	-0.48
HNR35	M	-8.72***	-13.35***	-4.63***
	F	-7.97***	-8.99***	-0.86

DECPA	M	-100.9***	-62.4***	38.5**
	F	-183.3***	-6.9	176.4***
CQ	M	-0.01***	-0.02***	-0.004
	F	-0.007	-0.0002	0.006

Although acoustic analysis has demonstrated that T11 is produced with breathy voice, the temporal characteristics of this non-modal phonation—whether it spans the entire vowel and where it primarily occurs—remain to be investigated. To examine this, the vowel of T11 is segmented into three intervals (onset: 01; mid: 02; offset: 03); moreover, the values at the three time points are calculated as averages across each time interval (Fig. 3). LMM results showed that the main effect of segment was significant for all parameters (ps < 0.05). Except for H1*-H2*, the interaction effect between segment and gender was also significant (ps < 0.05). Table 1 showed that, except for H1*-An* and CQ in females, most parameters differed significantly between males and females across different segments (ps < 0.05).

Fig. 3 is to be inserted here

As shown in Fig. 3 and Table 1, the first third of the vowel in males shows significantly higher values of H1*-H2* and H1*-A1* compared to the second and third intervals (ps < 0.01). H1*-A2* and H1*-A3* are also greater in the first interval than in the second interval, though not significantly different from the third interval (ps > 0.05). Overall, T11 has a steeper spectral tilt in the first third. CPP and HNRs are significantly lower in the first interval than in the other two intervals (ps < 0.001). Although significant differences are also observed between the second and third intervals for CPP and HNRs (ps < 0.05), these differences are considerably smaller than those between the first and second/third intervals. Similarly, CQ and DECPA values are significantly lower in the first interval compared to the other two intervals (ps < 0.05). For female speakers, a similar pattern is observed for most acoustic parameters except for H1*-An* and EGG measures.

Taken together, compared to the other two intervals, the acoustic and EGG measures of the first interval in T11—characterized by greater spectral tilt, increased noise, reduced glottal closure, and slower vocal fold closure—are generally consistent with the properties of breathy voice. These findings indicate that the non-modal phonation associated with T11 does not span the entire vowel but is localized primarily in the first third. Therefore, the subsequent analysis focuses on this interval to characterize fine-grained features of non-modal voices in T11.

3.2. Individual differences in non-modal phonation of T11

As illustrated in Fig. 1, substantial inter-speaker variability is observed in T11, particularly regarding its distinctive phonation, which predominantly occurs in the first third of the vowel. To facilitate the observation of the acoustic and physiological characteristics of the five level tones, as well as individual differences across the thirty speakers, data from the first third of the five level tones are extracted. These multidimensional acoustic and EGG measures are subjected to principal component analysis (PCA) (Gao and Kuang, 2020). Prior to the

PCA, for each parameter, the raw data (all tokens) for all participants were normalized using z-scores. This normalization procedure was implemented to eliminate scale effects arising from differences in measurement units and value ranges across parameters, thereby ensuring the comparability of variables in the PCA and preventing any specific parameter with a larger scale from exerting a disproportionate influence on component extraction. In PCA, following usual practice, variables with loadings (weights) of 0.32 or higher on a given component were considered to form a principal component (Tabachnick and Fidell, 2013; Lee et al., 2019).

Fig. 4 is to be inserted here

Fig. 4 indicate that the first two principal components account for 61% of the total variance (PC1: 46.1%; PC2: 14.9%). CPP and HNRs exhibit high negative loadings on PC1, while H1*-An* show strong positive loadings on PC1. Additionally, DECPA and CQ are also moderately negatively associated with PC1. This suggests that a larger PC1 value corresponds to a steeper spectral tilt, lower CPP and HNRs, and smaller CQ and DECPA. PC2 is most strongly correlated with H1*-H2* and H1*-An*, which exhibit a negative correlation with PC2, indicating that a higher PC2 value corresponds to smaller harmonic amplitude differences. Moreover, the vectors for CQ and H1*-H2*, as well as DECPA and H1*-An*, are located in opposite quadrants and oriented in roughly inverse directions, implying a potential negative correlation in the PCA space.

Fig. 5 is to be inserted here

In the acoustic space constructed by PC1 and PC2 (Fig. 5), T22, T33, T44, and T55 are closely clustered, whereas T11 is distinctly separated from the other level tones, trending toward the lower-right quadrant—particularly along the PC1 dimension. T11 shows higher PC1 values than the other level tones, indicating a higher noise level, reduced periodicity, and greater harmonic amplitude differences, overall reflecting breathy phonation. However, the distribution of T11 is more dispersed, showing considerable internal variability. Therefore, the data of T11 are selected to explore the phonation variation. Specifically, LMMs are constructed to examine the effects of age, gender, and individual differences between speakers on PC1 and PC2. The model formulas are as follows: PC1 ~ Gender * Age + (1|Speaker) + (1|CV) (Conditional R^2 = 0.774, Marginal R^2 = 0.128); PC2 ~ Gender + Age + (1|Speaker) + (1|CV) (Conditional R^2 = 0.722, Marginal R^2 = 0.191). The model results indicated a significant main effect of gender on PC1 (Gender: $F(1, 25.03) = 4.92$, $p = 0.036$) and PC2 (Gender: $F(1, 26.05) = 7.86$, $p = 0.009$), reflecting the notable acoustic and physiological differences between male and female speakers in T11. A significant interaction between gender and age is also found on PC1 (Age * Gender: $F(1, 25.02) = 5.64$, $p = 0.026$), indicating that the effect of age on PC1 differs by gender (Fig. 6). Specifically, PC1 values in males increase slightly with age and remain positive. In contrast, PC1 values in females decrease significantly with age, suggesting that older female speakers have a higher signal-to-noise ratio and a lower spectral tilt.

Fig. 6 is to be inserted here

Notably, the two models demonstrate relatively high conditional R^2 (PC1: 0.774; PC2: 0.722) but low marginal R^2 (PC1: 0.128; PC2: 0.191), indicating that random effects account for a substantial proportion of variance. To be specific, compared to fixed effects, random intercepts contribute more to variance in the dependent variables, explaining 64.6% (PC1 model) and 53.1% (PC2 model) of the variance, respectively. Table

2 lists variance estimates for speaker and CV. Among these, the speaker effect contributes the most to the random intercept variance in both models, indicating considerable individual differences among speakers in PC1 and PC2 within T11.

Table 2 Random effects of PC1 and PC2 in linear mixed effects models.

Variable	Groups	Name	Variance	Std. Dev.	Contribution (%)
PC1	Speaker	(Intercept)	1.97	1.40	70.5%
	CV	(Intercept)	0.10	0.32	3.6%
	Residual	-	1.001	0.85	25.9%
PC2	Speaker	(Intercept)	0.941	0.970	64.4%
	CV	(Intercept)	1.	1.17	1.2%
	Residual	-	0.502	0.709	34.3%

To explore these individual differences in phonations in T11, the random intercepts were used to screen for individuals with deviations from the group average (Fig. 7). Using the group average (i.e., breathy voice) as the reference, we screen the 30 speakers to identify those who exhibit acoustic patterns different from breathy voice. The specific screening criterion is: speakers with negative intercepts in the PC1 model but positive intercepts in the PC2 model. These speakers may use phonations other than the typical breathy voice to achieve phonemic contrasts between T11 and the other level tones. As shown in Fig. 7, eight speakers (M02, M05, M07, M09, M16, F01, F02, F04) met this criterion, suggesting relatively high signal-to-noise ratios and low spectral tilt (indicating a faster glottal closure rate) relative to the group average in T11 production.

Fig. 7 is to be inserted here

On this basis, the representative acoustic measures with high loadings on the first two principal components are selected to characterize the subtypes of phonation (Table 3). The parameter selection criteria include factor loadings greater than 0.32 and commonly used in cross-linguistic studies. Based on the two criteria, HNRs are chosen as the representative parameters for PC1, and H1*-H2* for PC2. Each speaker's T11 is compared with their T33 (modal reference, see Section 3.1) to investigate intra-speaker variation in phonation associated with T11. Besides, T11 has also been perceptually associated with harsh voice (Liu et al., 2020), thus, the articulation parameter (F1) is also incorporated into the analysis. A Wilcoxon signed-rank test is conducted on F1 (z-score normalized) for T11 and T33 across 30 native speakers. The results reveal no statistically significant difference between T11 and T33 for any speaker (ps > 0.05), although a slight increase in F1 for T11 relative to T33 is observed in a small subset of speakers. Fig. 8 compares the first third of T11 and T33 across the three acoustic parameters—HNRs, H1*-H2*, and F1—for the eight selected speakers (M02, M05, M07, M09, M16, F01, F02, F04).

Fig. 8 is to be inserted here

Table 3 Factor loadings of each parameter on the first two principal components (PC1, PC2).

·	PC1	PC2
H1*-H2*	0.16	**-0.49**
H1*-A1*	0.28	-0.43
H1*-A2*	0.30	-0.39
H1*-A3*	0.28	-0.31
CPP	-0.29	-0.10
HNR05	**-0.36**	-0.22
HNR15	**-0.39**	-0.27
HNR25	**-0.39**	-0.28
HNR35	**-0.38**	-0.30
CQ	-0.20	0.17
DECPA	-0.18	0.02

As shown in Fig. 8, three distinct patterns in the acoustic parameters are observed. First, speakers F01 and F04 exhibit no significant differences between T11 and T33 in H1*-H2*, HNRs, and F1 (ps > 0.05), indicating near-modal voice characteristics relative to the group average. Second, for speakers F02, M05, and M16, T11 shows a larger H1*-H2* ($p < 0.05$) and lower HNRs ($p < 0.05$) relative to T33, along with a significant increase in F1, resulting in a harsher voice overall. Third, speakers M02, M07, and M09 display no significant difference in H1*-H2* between T11 and T33 (ps > 0.05), but significant reductions in HNRs are observed for T11 compared to T33 (ps < 0.01). While no significant difference in F1 is found (ps > 0.05), an upward trend in F1 for T11 relative to T33 is evident, resulting in a voice quality approaching harshness.

3.3. Non-modal phonation and pitch range

By observing the pitch curves of the five level tones across 30 speakers (Fig. 1), some speakers exhibit a relatively large pitch range, allowing clear differentiation of the five level tones within the tonal space. However, others exhibit a limited pitch range, with F0 curves for T11 overlapping or closely approximating those for T22, indicating substantial individual variation. This raises the question of whether the number of cues (i.e., pitch, phonation) used by speakers in tonal distinctions is related to their pitch range, namely, whether an interaction exists between pitch range and phonation. Given this, the pitch range of each speaker was calculated as follows: the mean F0 of each token among the five level tones was computed; then, the difference between the maximum and minimum F0 values among these averages was selected as the pitch range of a specific speaker (Fig. 9).

Fig. 9 is to be inserted here

To investigate whether the use of non-modal phonations in T11 is associated with pitch range, Pearson's correlation coefficient is employed to examine the correlation between PC1 (random intercepts of speakers), PC2 (random intercepts of speakers), and the speakers' pitch range. The random intercepts of PC1 exhibit a statistically significant moderate negative correlation with pitch range (r = -0.435, p = 0.02); similarly, the random intercepts of PC2 show a statistically significant moderate positive correlation with pitch range (r = 0.409, p = 0.03). This suggests that as pitch range increases, the random intercepts for speakers in the PC1 model decrease, while those in the PC2 model increase. Considering the parameters associated with PC1 and PC2, it is found that an increase in spectral tilt and a decrease in the signal-to-noise ratio usually correspond to a reduction in pitch range. In other words, when producing the five level tones, speakers with a narrower pitch range are more inclined to employ non-modal phonations. However, it is important to note that, given the magnitude of the correlation coefficients, the relationship between pitch range and the use of non-modal phonations is correlational rather than causal.

4. Discussion

Within the framework of Dispersion Theory, this study presents a comprehensive acoustic and EGG analysis of the five level tones in Hmu based on data from 30 native speakers. Particular attention is given to the phonation and temporal characteristics of T11, and individual variation in its realization of multidimensional tonal distinction strategies. The main findings are as follows. First, the phonations of the five level tones have been defined strictly. T11is breathy, T55 is high-pitched, and the other level tones (T22, T33, and T44) are produced with modal voice. Meanwhile, the non-modal phonation associated with T11 is concentrated in the initial third of the syllable. Second, individual variation is observed in the phonation strategies to realize T11. In addition to breathy voice, certain speakers employ harsh or near-modal phonation. Third, the use of non-modal phonation is correlated with the pitch range of the five level tones, and the pitch range is at least one factor influencing the number of phonation cues activated. Speakers with narrower pitch ranges tend to rely more on breathy or harsh voice to maintain tonal contrasts.

Based on these findings, this section discusses the criteria for classifying phonations in the five level tones (Section 4.1) and the interaction between pitch and phonation (Section 4.2).

4.1 Phonation categories of the five level tones

Given that phonations exhibit cross-ethnic and cross-linguistic differences (Section 1.1), it is essential to establish a language-specific standard when defining phonation categories in Hmu. Notably, the term "modal voice" can be understood either as a contrastive phonation category relative to non-modal phonations or as a reference to a speaker's normal quality (Garellek, 2019). In Hmu, T33 is the tone produced by speakers in natural and comfortable conditions, and its acoustic and EGG parameters are centrally distributed across the five level tones (Section 3.1). Therefore, the phonation of T33 is defined as modal phonation, and the specific normative

data of T33 are shown in Table S3.

When establishing the reference values for modal phonation, we need to define the phonations of the other four level tones. As shown in Fig. 2, the acoustic and EGG parameters of T22 and T44 exhibit distributions largely consistent with those of T33, and no significant statistical differences are observed between them across most parameters (Table S2). These results indicate that the phonations of T22 and T44 align with those of T33, and thus classify them as modal phonation. Compared to T22, T33, and T44, T55 is characterized by higher F0, lower H1*-An*, and higher HNRs and CPP. Previous studies have described similar voice qualities as high-pitched voice (Kong, 2001) or tense voice (Kuang, 2013). Usually, tense voice is constricted, high-pitched, and regular-pitched, which means we would not necessarily expect a decrease in HNR relative to modal voice. Similarly, these characteristics are also observed in T55; however, its classification as high-pitched phonation rather than tense voice is based on recent cross-linguistic research (Keating et al., 2023): tense and lax voice are encompassed within the modal phonation continuum in the voice space, indicating that the primary distinction between tense and modal voice lies in pitch rather than phonation. Therefore, the term "high-pitched phonation" is adopted for T55.

As to T11, it has higher spectral tilt (H1*-H2* and H1*-An*) and lower HNRs and CPP, caused by vocal fold spreading during voicing (Klatt and Klatt, 1990; Simpson, 2012; Garellek et al., 2016). Additionally, its acoustic and EGG parameters differ significantly from those of the other level tones (Table S2), aligning with the phonatory characteristics of breathy voice. However, it is important to note that phonation variations in T11 are observed within the syllable and across speakers (Section 3). The distinctive phonation of T11 is concentrated in the first third of the vowel. Acoustic and EGG parameters in the first third differ significantly from those in the latter two-thirds, where a lower spectral tilt and higher HNRs and CPP are observed, showing no significant difference from T33. This suggests that the phonation of T11 is complex, with the first third exhibiting non-modal phonation, while the latter two-thirds conform to modal phonation.

Regarding inter-speaker variation of T11 in realizing multidimensional tonal distinction strategies, at least three distinct non-modal phonations are identified in the first third of T11 across individuals: breathy, harsh, and near-modal voice (Section 3.2). Usually, the harsh vowels are pharyngealized (higher F1) but have breathy voice initially during the vowel; namely, the vocal folds are more spread while the epilarynx is constricted (e.g., !Xóõ: Traill, 1986; Garellek, 2020; Keating et al., 2023). This phenomenon is also observed in certain speakers when producing T11 in Hmu. Among the speakers selected based on the random effects of LMM, an upward trend in F1 is observed, although this did not reach statistical significance (Fig. 8). Actually, the breathy and harsh categories cluster together in the voice space, indicating greater similarity between breathy voice and harsh voice (Keating et al., 2023). Therefore, the three phonation subtypes in T11 should not be interpreted as discrete phonological categories, but as alternative phonetic strategies available to speakers for enhancing tonal contrast.

Based on these intra-vowel and inter-speaker phonation patterns, T11 can be categorized into three phonation subtypes with complex clusters: breathy-modal voice, harsh-modal voice, and near-modal-modal

voice. This indicates that non-modal phonation is most prominent during the initial third of the vowel of T11. Notably, the subtypes of the non-modal phonations on T11 in Hmu are based on individual variations within the same language, differing from previous classifications that distinguished between creaky voice (prototypical creak, vocal fry, multiply pulsed voice, aperiodic voice, non-constricted creak, and tense/pressed voice; Keating et al., 2015) and breathy voice (slack/lax voice in Southern Yi, whispery voice in Shanghainese Wu, and [true] breathy voice in Gujarati and White Hmong; Tian and Kuang, 2021) based on cross-linguistic phonation variations. This might imply that phonation variations among individuals within the same language are no less than those across languages.

Additionally, languages with complex clusters of phonations include but are not limited to !Xóõ (breathy-creaky: Ladefoged, 1983; Traill, 1985, 1986) and Takhian Thong Chong (breathy-tense: DiCanio, 2009). On this basis, the phonation and articulation parameters are used together to describe the different phonation characteristics of the five level tones in Hmu. Therefore, in addition to the commonly used spectral tilt and noise considerations when defining phonation, articulation parameters also need to be considered, especially when defining non-modal phonations such as harsh voice that involve supraglottal constriction.

These findings extend Dispersion Theory by suggesting that dispersion operates not only at the level of phonological inventories but also at the level of individual cue selection within a shared system. Based on data from 30 speakers, we reveal individual differences in how speakers utilize phonation cues within the same speech community. When producing the same tone (T11), speakers exhibit diverse phonatory realizations, including breathy, harsh, and near-modal voice. Regardless of the phonatory strategy employed, the phonemic contrast between T11 and the other level tones is consistently maintained, highlighting that phonetic realizations must be understood in relation to their functional role in establishing linguistic contrasts. While individual variability may manifest in complex acoustic patterns, such variation is not random but ordered heterogeneity (Weinreich, 1968). Furthermore, the existence of these variations reflects the richness and dynamism of language, and the tolerance of internal variation leaves room and possibilities for language change. This study thus provides a referential analysis method for future investigations into individual variation in phonation. Research on the group characteristics and individual variation of phonation may contribute to a broader understanding of the physiological and sociolinguistic mechanisms underlying speech production.

4.2 Interaction between phonation and pitch

Regarding the relationship between pitch and phonation, it is also the case that languages may differ in the pitch ranges and voice qualities they typically use (Esling et al., 2019). Numerous languages utilize pitch, phonation, or both as phonemic distinctions in tonal languages. However, the interactive relationship between pitch and phonation in distinguishing phonological contrasts has been rarely explored. The complex tonal system of Hmu, which comprises five level tones and includes non-modal phonations, makes it an ideal candidate for investigating this interaction.

Given the variations in pitch range among the five level tones across 30 speakers, particularly the overlap between T11 and T22 in F0 (Fig. 1), this raises an important question: how is T11 distinguished from the other level tones in the crowded tonal system? Dispersion Theory provides a framework for addressing this issue. To maintain sufficient perceptual contrast while preserving the systematicity of the phonological structure, each pair of tones must exhibit at least one feature that allows speakers to easily differentiate them, while still permitting variation within individual tones. In essence, normalization is the process of mapping various acoustic realizations onto a single tonal category. Each tone occupies a distinct acoustic space, and to ensure a balance between perceptual clarity and articulatory economy, acoustic space is inherently variable.

Acoustic analysis of data from 30 speakers has revealed that the other four level tones (T22, T33, T44, T55) are primarily distinguished by pitch height. Although T55 exhibits high-pitched phonation, previous discussions suggest that high-pitched phonation falls within the range of modal phonation (Section 4.1). Consequently, these four level tones are sufficiently differentiated by pitch alone. The distinction between T11 and the other level tones, particularly T22, follows three patterns: (1) T11 and T22 are fully distinguished by pitch alone (e.g., F04, F06, M02, M09); (2) T11 and T22 are not well-distinguished in the acoustic space, exhibiting a narrow pitch range with overlapping F0 contours (e.g., F05, F09, F10); (3) T11 and T22 show slight separation at onset but converge in pitch contour during the latter half of the syllable (e.g., M08, M13). For the latter two cases, it is observed that speakers with a smaller pitch range are more likely to employ non-modal phonation when producing T11 to enhance its differentiation from T22. In other words, the phonemic contrast between T11 and T22 relies not solely on pitch but also incorporates phonation as a secondary cue. Perceptual study also indicates that pitch (primary cue) and phonation (secondary cue) contribute to the distinction between T11 and T22 in Hmu (Liu et al., 2024).

This study provides further empirical evidence that tonal contrast is multidimensional and that both single-cue and multi-cue differentiation strategies may be employed in tonal distinctions. In languages that rely on a single cue—such as pitch—for tonal distinction, two primary strategies may be adopted when the number of tones with the same contour increases, as in the case of five level tones in Hmu. One is to expand the pitch range, thereby maximizing tonal spacing within a single-dimensional pitch space (e.g., F01 and F04). The other involves using multiple cues, combining pitch with phonation (e.g., M03 and M20). These findings suggest that Hmu speakers employ different differentiation strategies for T11 and T22 to maximize phonemic contrast, thereby preventing phonemic mergers and ensuring communicative efficiency. This study offers empirical support for Dispersion Theory, which posits that a trade-off between perceptual distinctiveness and articulatory economy regulates the tonal system.

Furthermore, the five level tones provide an ideal case for exploring the interaction between pitch and phonation. According to the Multi-Register and Four-Level tonal model (Zhu, 2012), each register is divided into no more than four levels. However, in Hmu, F01 and F04 were found to produce T11 with near-modal voice. Both speakers have relatively wide pitch ranges among females, suggesting that when a speaker's pitch range is sufficiently broad to accommodate clear pitch distinctions among the five level tones, a single phonation category (i.e., modal voice) may suffice. This phenomenon is consistent with observations of the five level tones in Ziyun Miao (Kong, 1992). The statistical results of this study reveal a correlation between the speaker's pitch range and the use of non-modal phonations. Although causality cannot be established, the results at least suggest that the number of acoustic cues employed may be constrained by an individual's pitch range. Usually, the wider pitch ranges are associated with fewer cues; conversely, the narrower pitch ranges are associated with more cues.

Moreover, individual variation in tonal distinction may arise from the interaction of universal communicative pressures (contrast maintenance), language-specific structural conditions (tonal crowding), and individual physiological differences.

5. Conclusion

Phonation has garnered substantial attention, revealing various characteristics of laryngeal phonatory settings in tonal distinction. However, relatively little research has examined how individuals within the same speech community employ multidimensional cues to achieve tonal contrasts. The five level tones in Hmu provide a valuable case for exploring this issue. In this study, the three middle-level tones (T22, T33, and T44) exhibit characteristics of modal voice. T55 is characterized by a high-pitched voice due to its high F0. These four tones can therefore be distinguished solely by pitch. To maintain sufficient distinction in the crowded phonetic space, T11 achieves its contrast by introducing an additional phonation dimension (i.e., breathy voice), and this non-modal phonation is mainly concentrated in the first third of the vowel. At the individual level, considerable inter-speaker phonation variation is observed in the realization of T11, including breathy, harsh, and near-modal voice. Furthermore, the use of non-modal phonation correlates with the pitch range of the five level tones: speakers with wider pitch ranges tend to rely on pitch, whereas those with narrower pitch ranges are more likely to employ non-modal phonation cue. These findings suggest that speakers within the same speech community may adopt divergent yet equally effective strategies for achieving phonemic contrasts. Overall, this study not only provides an analytical method for exploring individual variability within the tonal system but also offers a preliminary reference for the relationship between multidimensional tonal distinction and individual realization in tonal typology.

Data Availability

The data pertinent to this study can be found in the Supplementary Information section. For detailed raw data, interested teams may contact the corresponding author with reasonable requests.

References

Andruski JE, Ratliff M (2000) Phonation types in production of phonological tone: the case of Green Mong. J Int Phon Assoc 30:37–61. https://doi.org/10.1017/S0025100300006654

Baken RJ, Orlikoff RF (2000) Clinical measurement of speech and voice. Singular Publishing Group, San Diego

Becker-Kristal R (2010) Acoustic typology of vowel inventories and dispersion theory: Insights from a large cross-linguistic corpus. University of California Press, Los Angeles

Bishop J, Keating P (2012) Perception of pitch location within a speaker's range: Fundamental frequency, voice quality and speaker sex. J Acoust Soc Am 132:1100–1112. https://doi.org/10.1121/1.4714351

Blankenship B (1997) The time course of breathiness and laryngealization in vowels. Dissertation, University of California

Blankenship B (2002) The timing of nonmodal phonation in vowels. J Phon 30:163–191. https://doi.org/10.1006/jpho.2001.0155

Brunelle M, Kirby J (2016) Tone and phonation in Southeast Asian languages. Lang Linguist Compass 10:191–207. https://doi.org/10.1111/lnc3.12182

Chai Y, Garellek M (2022) On H1−H2 as an acoustic measure of linguistic phonation type. J Acoust Soc Am 152:1856–1870 https://doi.org/10.1121/10.0014175

Chao Y (1948) Mandarin primer: An intensive course in spoken Chinese. Harvard University Press, Cambridge

Chen C, Havenhill J (2025) Harsh voice and its interaction with vowel quality in Fuzhou Min Chinese. J Acoust Soc Am 157:2582–2602 https://doi.org/10.1121/10.0036256

de Krom G (1993) A cepstrum-based technique for determining harmonics-to-noise ratio in speech signals. J Speech Lang Hear Res 36:254–266. https://doi.org/10.1044/jshr.3602.254

Denning K (1989) The diachronic development of phonological voice quality, with special reference to Dinka and the other Nilotic languages. Dissertation, Stanford University

DiCanio CT (2009) The phonetics of register in Takhian Thong Chong. J Int Phon Assoc 39:162–188. https://doi.org/10.1017/S0025100309003879

Edmondson JA, Esling JH (2006) The valves of the throat and their functioning in tone, vocal register and stress: laryngoscopic case studies. Phonology 23:157–191. https://doi.org/101017/S095267570600087X

Esling JH, Harris JG (2005) States of the glottis: An articulatory phonetic model based on laryngoscopic observations. In: Hardcastle WJ, Beck JM (eds) A figure of speech: A festschrift for John Laver. Erlbaum, Mahwah, p 347–383

Esling JH, Moisik SR, Benner A, Crevier-Buchman L (2019) Voice quality: The laryngeal articulator model. Cambridge University Press, Cambridge

Esposito CM (2012) An acoustic and electroglottographic study of White Hmong tone and phonation. J Phon 40:466–476. https://doi.org/10.1016/j.wocn.2012.02.007

Esposito CM, Khan SD (2020) The cross-linguistic patterns of phonation types. Lang Linguist Compass 14:e12392 https://doi.org/10.1111/lnc3.12392

Fabre P (1957) Un procede electrique d inscription de I accolement glottique aucours de laphonation: glottographie de haute frequence. Bulletin de l'Académie nationale de médecine, 141, 66-69.

Flemming E (1995) Audio representations in phonology. Dissertation, University of California.

Fourcin AJ, Abberton E (1971) First applications of a new laryngograph. Medical and Biological Illustration 21:172–182.

Gao X, Kuang J (2022) Phonation Variation as a Function of Checked Syllables and Prosodic Boundaries. Language 7(3):171 https://doi.org/10.3390/languages7030171

Garellek M (2019) The phonetics of voice. In: Katz WF, Assmann PF (eds) Routledge handbook of phonetics. Routledge, Oxford, p 75–106

Garellek M (2020) Acoustic discriminability of the complex phonation system in !Xóõ. Phonetica 77:131–160. https://doi.org/10.1159/000494301

Garellek M (2022) Theoretical achievements of phonetics in the 21st century: Phonetics of voice quality. J Phon 94:101155 https://doi.org/10.1016/j.wocn.2022.101155

Garellek M, Chai Y, Huang Y, Van Doren M (2021) Voicing of glottal consonants and non-modal vowels. J Int Phon Assoc 53:305–332. https://doi.org/10.1017/S0025100321000116

Garellek M, Keating P (2011) The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. J Int Phon Assoc 41:185–205. https://doi.org/10.1017/S0025100311000193

Garellek M, Keating P, Esposito CM, Kreiman J (2013) Voice quality and tone identification in White Hmong. J Acoust Soc Am 133:1078–1089. https://doi.org/10.1121/1.4773259

Garellek M, Ritchart A, Kuang J (2016) Breathy voice during nasality: A cross-linguistic study. J Phon 59:110–121 https://doi.org/10.1016/j.wocn.2016.09.001

Gerfen C, Baker K (2005) The production and perception of laryngealized vowels in Coatzospan Mixtec. J Phon 33:311–334. https://doi.org/10.1016/j.wocn.2004.11.002

Gerratt BR, Kreiman J (2001) Toward a taxonomy of nonmodal phonation. J Phon 29:365–381 https://doi.org/10.1006/jpho.2001.0149

Gordon M, Ladefoged P (2001) Phonation types: a cross-linguistic overview. J Phon 29:383–406. https://doi.org/10.1006/jpho.2001.0147

Hillenbrand J, Cleveland RA, Erickson RL (1994) Acoustic correlates of breathy vocal quality. J Speech Lang Hear Res 37:769–778. https://doi.org/10.1044/jshr.3704.769

Holmberg EB, Hillman RE, Perkell JS, Guiod P, Goldman SL (1995) Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice. J Speech Lang Hear Res 38:1212–1223. https://doi.org/10.1044/jshr.3806.1212

Howard DM (1995) Variation of electrolaryngographically derived closed quotient for trained and untrained adult female singers.J Voice 9:163–172. https://doi.org/10.1016/S0892-1997(05)80250-4

Huffman MK (1987) Measures of phonation types in Hmong. J Acoust Soc Am 81:495–504 https://doi.org/10.1121/1.394915

Iseli M, Shue Y, Alwan A (2007) Age, sex, and vowel dependencies of acoustical measures related to the voice source. J Acoust Soc Am 121:2283–2295. https://doi.org/10.1121/1.2697522

Jessen M, Roux JC (2002) Voice quality differences associated with stops and clicks in Xhosa. J Phon 30:1–52. https://doi.org/10.1006/jpho.2001.0150

Keating P, Esposito C, Garellek M, Khan SD, Kuang J (2011) Phonation contrasts across languages. Paper presented at the Proceedings of the 17th International Congress of Phonetic Sciences, ICPhS, Hong Kong, 1046–1049

Keating P, Esposito CM, Garellek M, Khan SD, Kuang J (2010) Phonation contrasts across languages. UCLA Working Papers in Phonetics 108:188–202

Keating P, Garellek M, Kreiman J (2015) Acoustic properties of different kinds of creaky voice. Paper presented at the Proceedings of the 18th International Congress of Phonetic Sciences, ICPhS, Glasgow, (2–7

Keating P, Kuang J, Garellek M, Esposito CM, Khan SD (2023) A cross-language acoustic space for vocalic phonation distinctions. Language 99:351–389. https://doi.org /10.1353/lan.2023.a900090

Klatt DH, Klatt LC (1990) Analysis, synthesis, and perception of voice quality variations among female and male talkers. J Acoust Soc Am 87:820–857. https://doi.org/10.1121/1.398894

Kong J (1992) Acoustic and perceptual studies on the five-level tones of Ziyun Miao. In XMa JWang (Eds.), A new preliminary study of ethnic languages, Sichuan Ethnic Publishing House, Chengdu, 152–163

Kong J (2001) On language phonation. Central University for Nationalities Press, Beijing

Kreiman J, Gerratt BR, Antoñanzas-Barroso N (2007) Measures of the glottal source spectrum. J Speech Lang Hear Res 50:595–610. https://doi.org/10.1044/1092-4388(2007/042)

Kreiman J, Gerratt BR, Garellek M, Samlan R, Zhang Z (2014) Toward a unified theory of voice production and perception. Loquens 1:e009. https://doi.org/10.3989/loquens.2014.009

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

Kuang J, Cui A (2018) Relative cue weighting in production and perception of an ongoing sound change in Southern Yi. J Phon 71:194–214. https://doi.org/10.1016/j.wocn.2018.09.002

Kuang J, Keating P (2014) Vocal fold vibratory patterns in tense versus lax phonation contrasts. J Acoust Soc Am 136:2784–2797. https://doi.org/10.1121/1.4896462

Kuznetsova A, Brockhoff PB, Christensen RHB (2017) lmerTest Package: Tests in Linear Mixed Effects Models. J Stat Softw 82:1–26. https://doi.org/10.18637/jss.v082.i13

Ladefoged P (1971) Preliminaries to linguistic phonetics. University of Chicago, Chicago

Ladefoged P (1983) The linguistic use of different phonation types. In DBless JAbbs (Eds.), Vocal fold physiology: Contemporary research and clinical issues. College-Hill Press, San Diego CA

Ladefoged P, Maddieson I (1996) The Sounds of the World's Languages. Blackwell, Oxford

Laver J (1980) The phonetic description of voice quality. Cambridge University Press, Cambridge

Lee Y, Keating P, Kreiman J (2019) Acoustic voice variation within and between speakers. J Acoust Soc Am 146:1568–1579. https://doi.org/10.1121/1.5125134

Liljencrants J, Lindblom B (1972) Numerical simulation of vowel quality systems: The role of perceptual contrast. Language 48:839-862. https://doi.org/10.2307/411991

Lindblom B (1986) Phonetic universals in vowel systems. In JJOhala JJJaeger (Eds.), Experimental phonology, Academic Press, 13–44

Lindblom B (1990) Phonetic content in phonology. Phonetic Experimental Research at the Institute of Linguistics University of Stockholm, 11:100–118

Lindblom B, Maddieson I (1988) Phonetic universals in consonant systems. In LHyman CNLi (Eds.), Language, speech, and mind: Studies in honour of Victoria AFromkin, Routledge, 62–78

Liu W (2020) A perceptual study on the five level tones in Hmu (Xinzhai variety). Paper presented at the Proceedings of Interspeech 2020, Shanghai, China, 1620–1623

Liu W (2021) Physiological and physical basis of voice quality and its linguistic value. Essays on linguistics 63:204-233

Liu W, Lin Y-J Yang Z, Kong J (2020) Hmu (Xinzhai variety). J Int Phon Assoc 50:240–257. https://doi.org/10.1017/S0025100318000336

Liu W, Peng G, Kong J (2024) The role of breathy voice in Hmu tone perception. J Chin Linguist 52:138–174. https://doi.org/10.1353/jcl.2024.a919401

Liu W, Wang F, Kong J (2019) An acoustic study on the phonation variations of tones in Bai (Beiwuliqiao variety). Contemporary Linguistics 1:119–138

Maddieson I (1978) Universals of tone. In JHGreenberg CAFerguson, EAMoravcsik (Eds.), Universals of human language: Vol. 2. Phonology, Stanford University Press, Stanford CA, 335–365

Maddieson I, Ladefoged P (1985) Tense and lax in four minority languages of China. J Phon 13:433–454. https://doi.org/10.1016/S0095-4470(19)30788-0

Miller AL (2007) Guttural vowels and guttural co-articulation in Ju|'hoansi. J Phon 35:56–84. https://doi.org/10.1016/j.wocn.2005.11.001

Moisik SR (2013) Harsh voice quality and its association with blackness in popular american media. Phonetica 69:193–215. https://doi.org/10.1159/000351059

Moisik SR, Czaykowska-Higgins E, Esling JH (2021) Phonological potentials and the lower vocal tract. J Int Phon Assoc 51:1–35. https://doi.org/10.1017/S0025100318000403

Moisik SR, Esling JH (2014) Modeling the biomechanical influence of epilaryngeal stricture on the vocal folds: A low-dimensional model of vocal–ventricular fold coupling. J Speech Lang Hear Res 57:S687–S704. https://doi.org/10.1044/2014_JSLHR-S-12-0279

Moisik SR, Esling JH, (2011) The 'whole larynx' approach to laryngeal features. In Proceedings of the 17th International Congress of Phonetic Sciences, ICPhS, Hong Kong, 1406–1409

Moisik SR, Lin H, Esling JH (2014) A study of laryngeal gestures in Mandarin citation tones using simultaneous laryngoscopy and laryngeal ultrasound (SLLUS). J Int Phon Assoc 44:21–58. https://doi.org/10.1017/S0025100313000327

Rothenberg M, Mahshie JJ (1988) Monitoring vocal fold abduction through vocal fold contact area. J Speech Lang Hear Res 31:338–351. https://doi.org/10.1044/jshr.3103.338

Schwartz J-L Boë, L-J Vallée N, Abry C (1997a) Major trends in vowel system inventories. J Phon 25:233–253. https://doi.org/10.1006/jpho.1997.0044

Schwartz J-L Boë, L-J Vallée N, Abry C (1997b) The dispersion-focalization theory of vowel systems. J Phon 25:255–286. https://doi.org/10.1006/jpho.1997.0043

Shue Y-L Keating P, Vicenik C, Yu KM (2011) VoiceSauce: A program for voice analysis. In Proceedings of the 17th International Congress of Phonetic Sciences, ICPhS, Hong Kong, 1846-1849

Silverman D (1997) Phasing and recoverability. Garland Publishing, New York

Silverman D, Blankenship B, Kirk PL, Ladefoged P (1995) Phonetic structures in Jalapa Mazatec. Anthropol Linguist 37:70–88. https://www.jstor.org/stable/30028043

Simpson AP (2012) The first and second harmonics should not be used to measure breathiness in male and female voices. J Phon 40:477–490. https://doi.org/10.1016/j.wocn.2012.02.001

Stevens KN (1977) Physics of laryngeal behavior and larynx modes. Phonetica 34:264–279. https://doi.org/10.1159/000259885

Tạ, TT Brunelle M, Nguyễn TQ (2022) Voicing and register in Ngãi Giao Chrau: Production and perception studies. J Phon 90:101115. https://doi.org/10.1016/j.wocn.2021.101115

Tabachnick B, Fidell L (2013) Using multivariate statistics. Pearson Education Inc, Boston MA

Tehrani H (2009) EGGWorks: a program for automated analysis of EGG signals. http://www.linguistics.ucla.edu/faciliti/facilities/physiology/Egg WorksSetup.exeS

Tian J, Kuang J (2021) The phonetic properties of the non-modal phonation in Shanghainese. J Int Phon Assoc 51:202–228. https://doi.org/10.1017/S0025100319000148

Traill A (1985) Phonetic and phonological studies of the !Xóõ Bushmen. Hamburg, Buske

Traill A (1986) The laryngeal sphincter as a phonatory mechanism in !Xóõ Bushman. In RSinger JKLundy (Eds.), Variation, culture and evolution in African populations: Papers in honor of Dr. Hertha de Villiers, Witwatersrand University Press, Johannesburg, 123–131

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

Weinreich U, Labov W, Herzog M (1968) Empirical foundations for a theory of language change. University of

Texas Press, Austin
Zhu X (2012) Multiregisters and four levels: A new tonal model. J Chin Linguist 40:1–17. https://www.jstor.org/stable/23754196

Acknowledgments

This work is supported by National Social Science Foundation (No. 22CYY022) and Jiangsu Oral Culture Corpus Transcription Project (No. HXSK 2023003). We would like to give our thanks to all participants for their excellent help in Xinzhai village with the recording, especially for Zhenghui Yang for his assistance.

Author contributions

Wen Liu: Conceptualization, Methodology, Investigation, Data curation, Writing – original draft, Writing – review and editing, Funding acquisition. Nianhan Hou: Writing – original draft, Writing – review and editing, Formal analysis, Visualization. Hao Tang: Conceptualization, Writing – review and editing, Funding acquisition.

Competing interests

The authors declare no competing interests.

Ethical approval

This study was approved by the Ethics Committee of Shandong University (Approval No. SDU-2021-307) on December 10, 2021. All procedures involving human participants were conducted in accordance with the ethical standards of the Declaration of Helsinki.

Informed consent

This study obtained written informed consent from all participants on August 17, 2022, prior to their enrollment. All participants were clearly informed of the study's purpose, procedures, data usage methods, and participants' rights (including the principle of voluntary participation and the right to withdraw unconditionally at any time). Consent forms were documented in writing and collected directly from adult participants with independent legal capacity to consent. No personally identifiable information was recorded or disclosed in any data. This study did not involve any vulnerable populations, and all participants provided their consent voluntarily.

Fig. S1 Mean F0 curves of all tokens for each tone in Hmu across nine normalized time points (left: male speakers; right: female speakers).

Fig. 1 Mean F0 curves of all tokens for the five level tones produced by individual speakers (20 males, 10 females) in Hmu across nine normalized time points.

Fig. 2 Boxplots of acoustic and EGG parameters across the five level tones for males and females.

Fig. 3 Boxplots of acoustic and EGG measures across three intervals (onset: 01; mid: 02; offset: 03) in T11 for males and females.

Fig. 4 The loadings for PC1 and PC2 of all acoustic and EGG features among the five level tones.

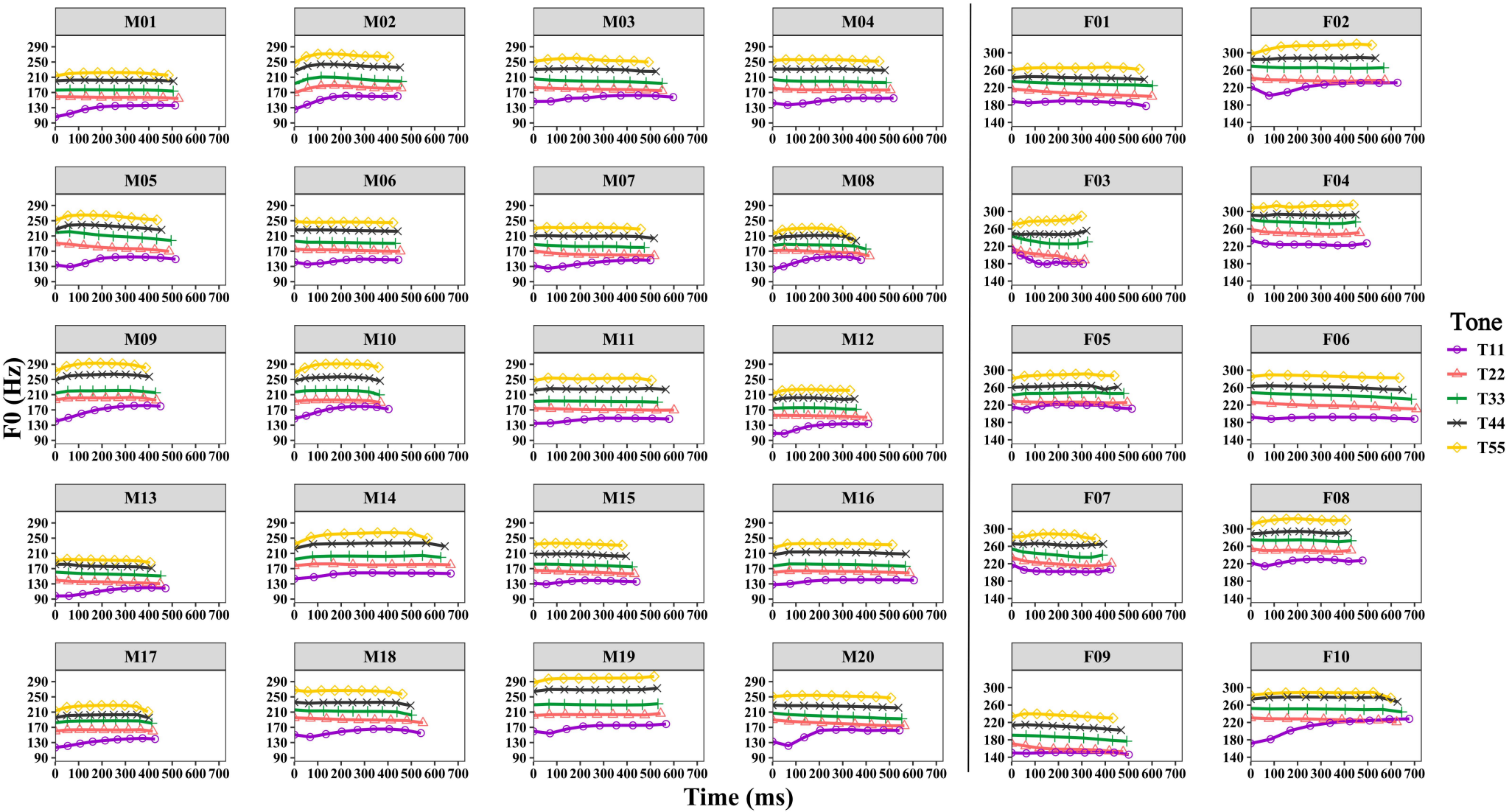
Fig. 5 PC1 and PC2 scores across all tokens among the five level tones (with 95% confidence ellipses).

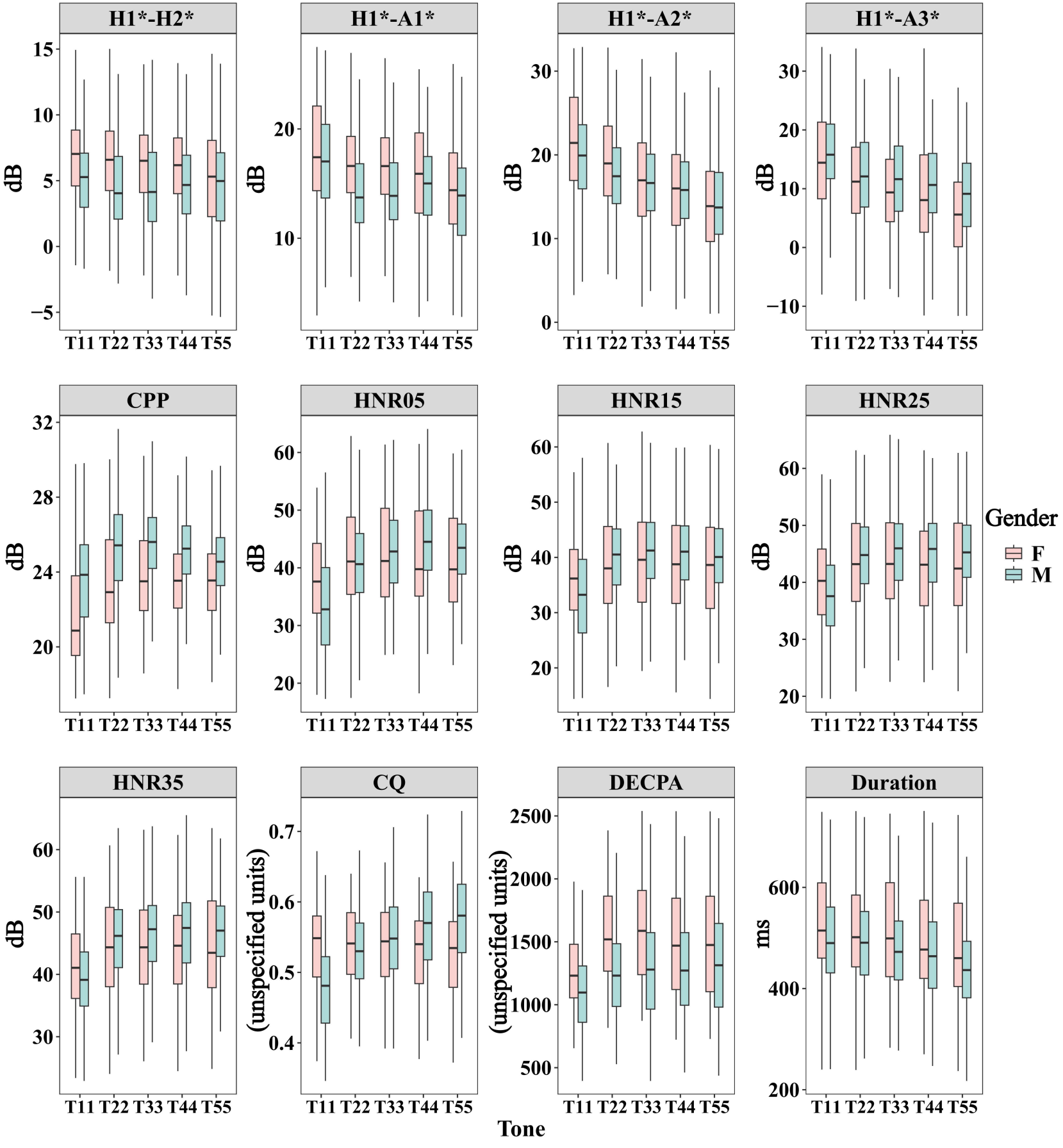
Fig. 6 Age*Gender interaction effect on PC1 in T11 (with 95% confidence intervals).

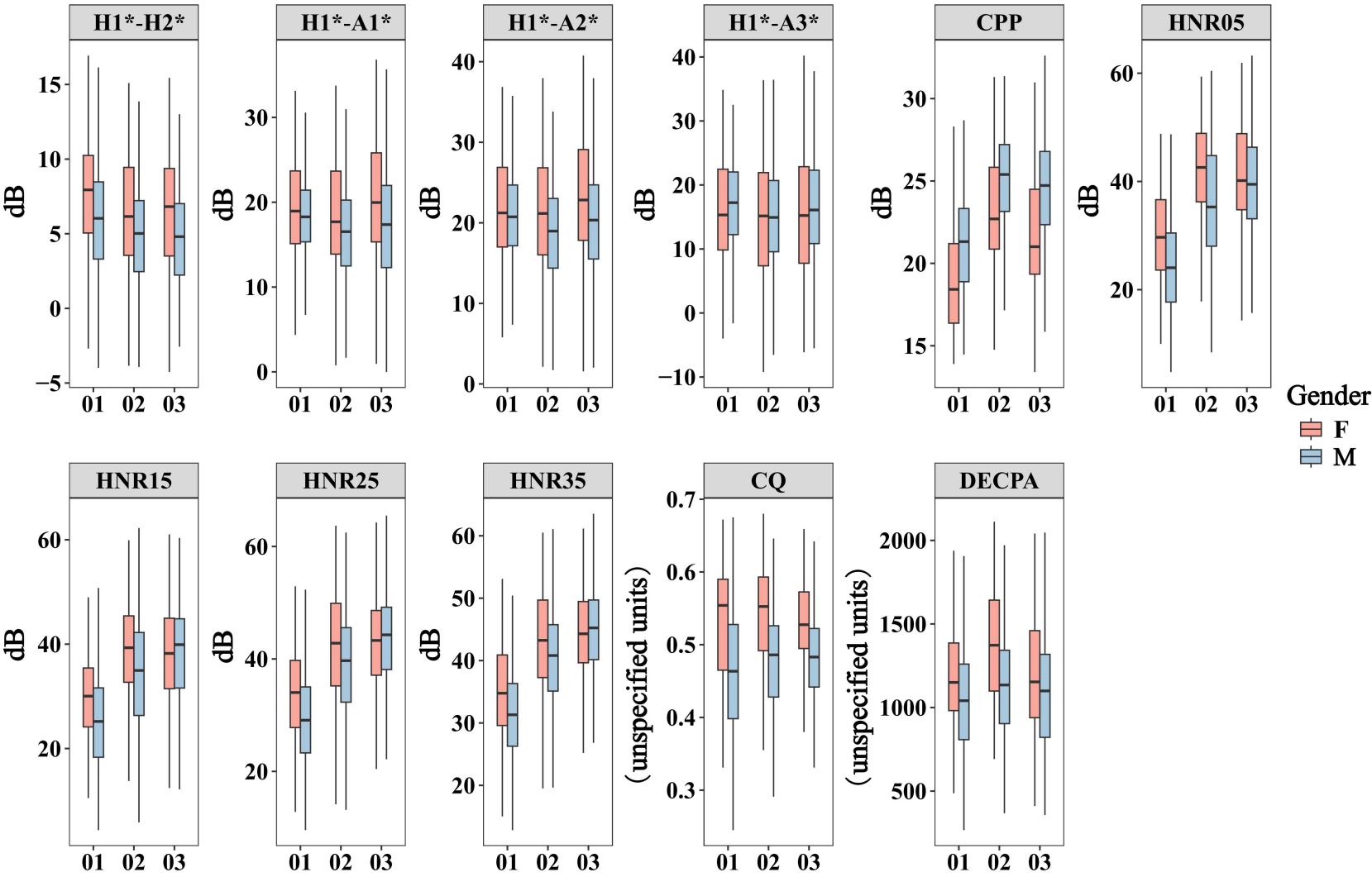
Fig. 7 Random intercept of PC1 model (upper panel) and PC2 model (lower panel) in T11 across 30 speakers. Bars near zero represent speakers whose PC1 and PC2 values approximate the group average; in contrast, bars deviating from zero represent speaker-specific deviations from the group average, with longer bars indicating greater deviations and a larger difference from the group average.

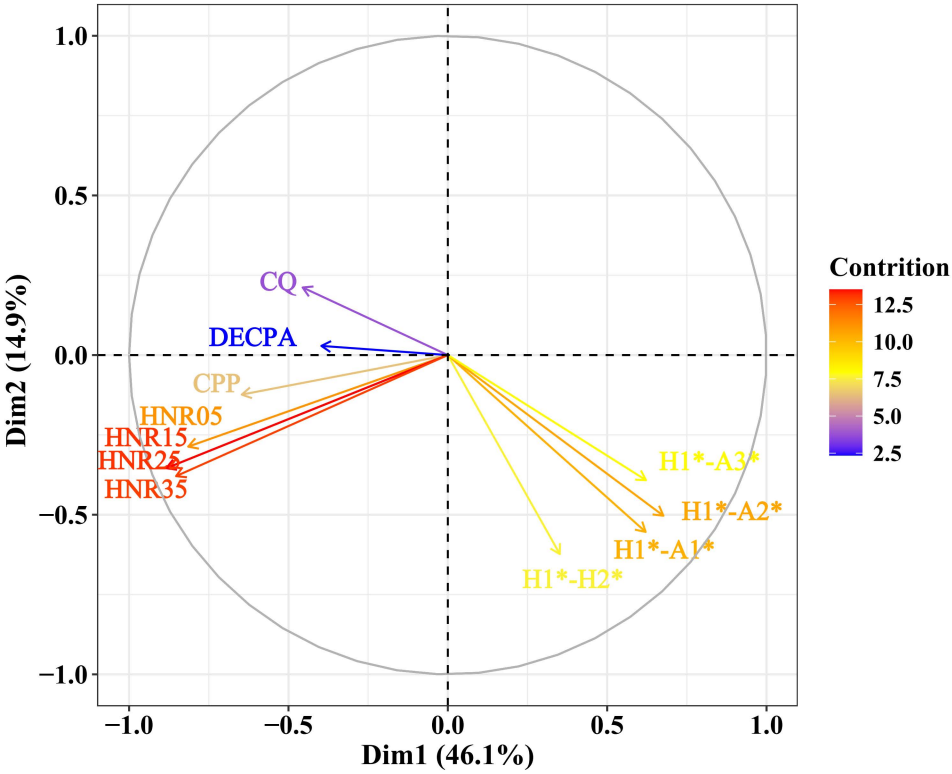
Fig. 8 Boxplots of H1*-H2*, HNRs, and F1 between T11 and T33 for eight speakers (** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ns = no significant, *p*-values corrected with Bonferroni).**

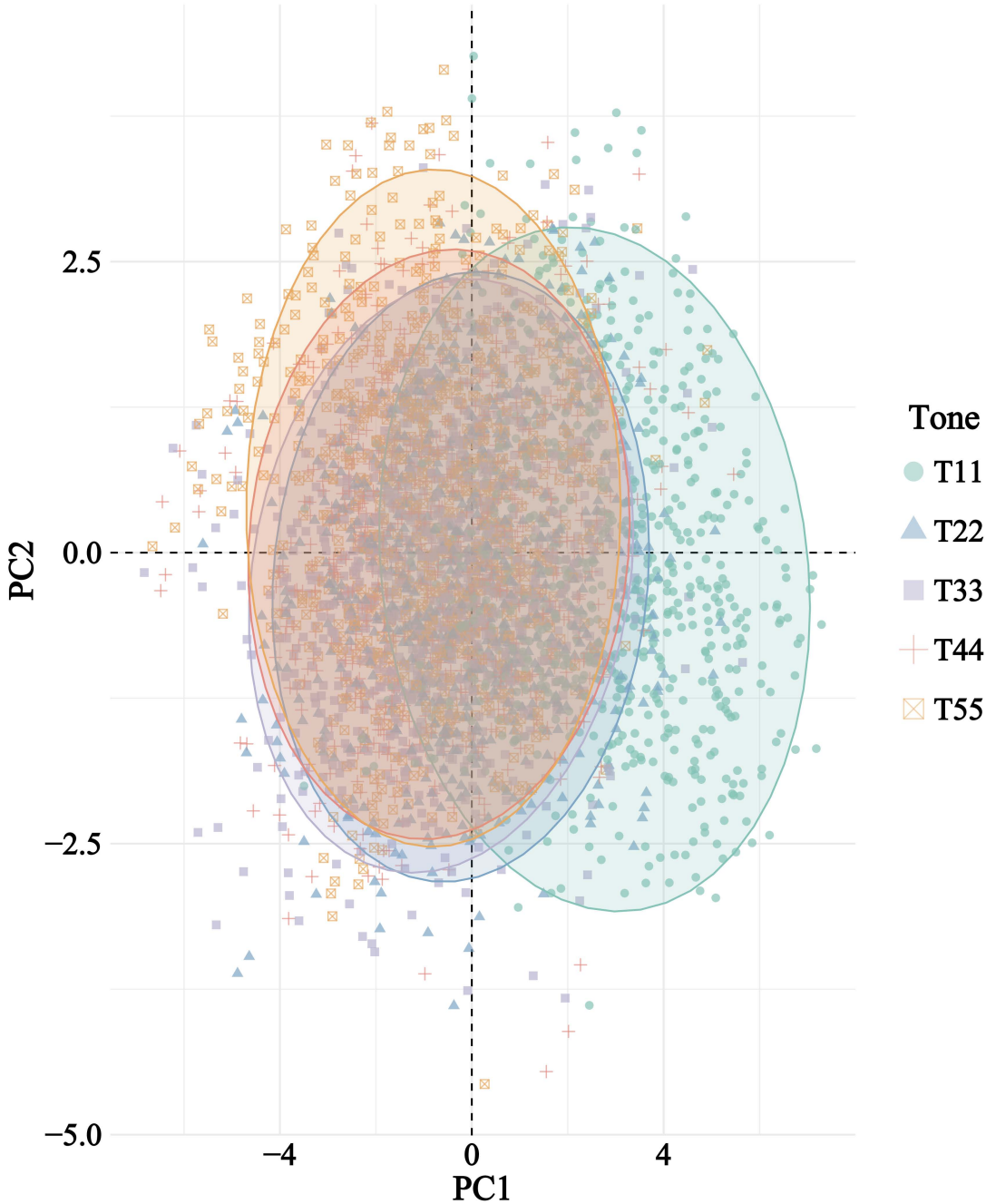
Fig. 9 Pitch range of the five level tones across thirty speakers in Hmu.

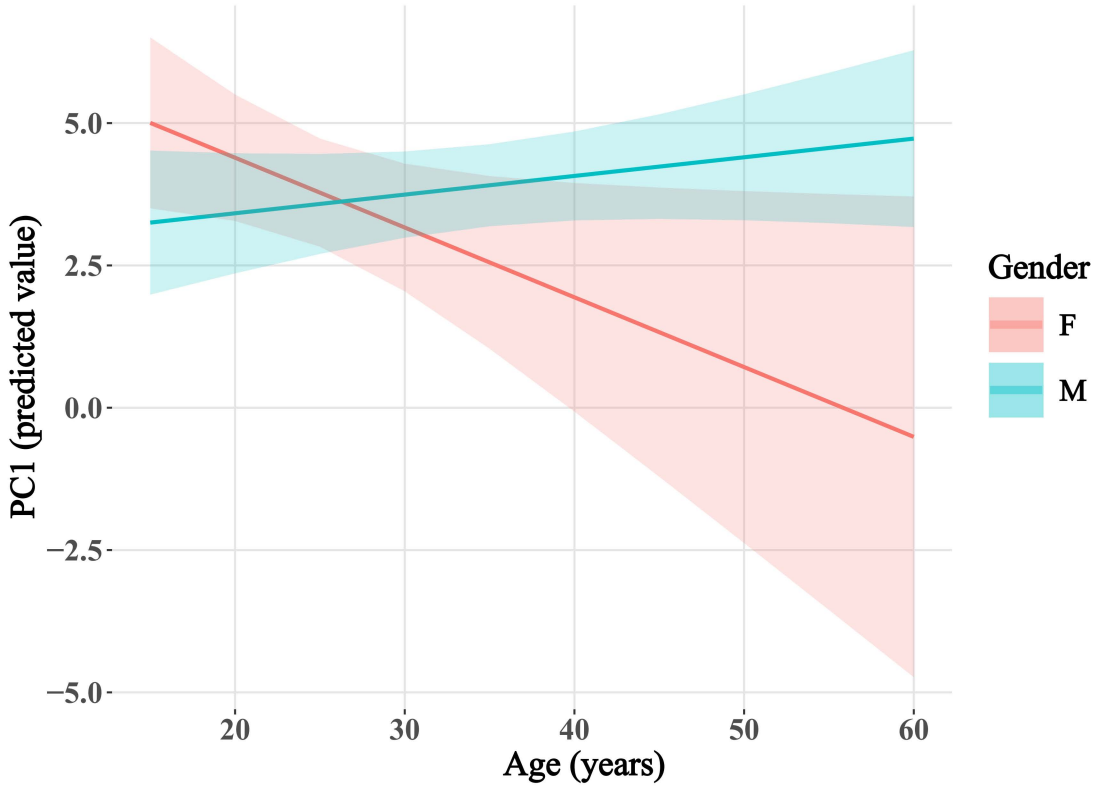

M01
M02
M03
M04
F01
F02
M05
M06
M07
M08
F03
F04
M09
M10
M11
M12
F05
F06
M13
M14
M15
M16
F07
F08
M17
M18
M19
M20
F09
F10
F0 (Hz)
Time (ms)
Tone
T11
T22
T33
T44
T55

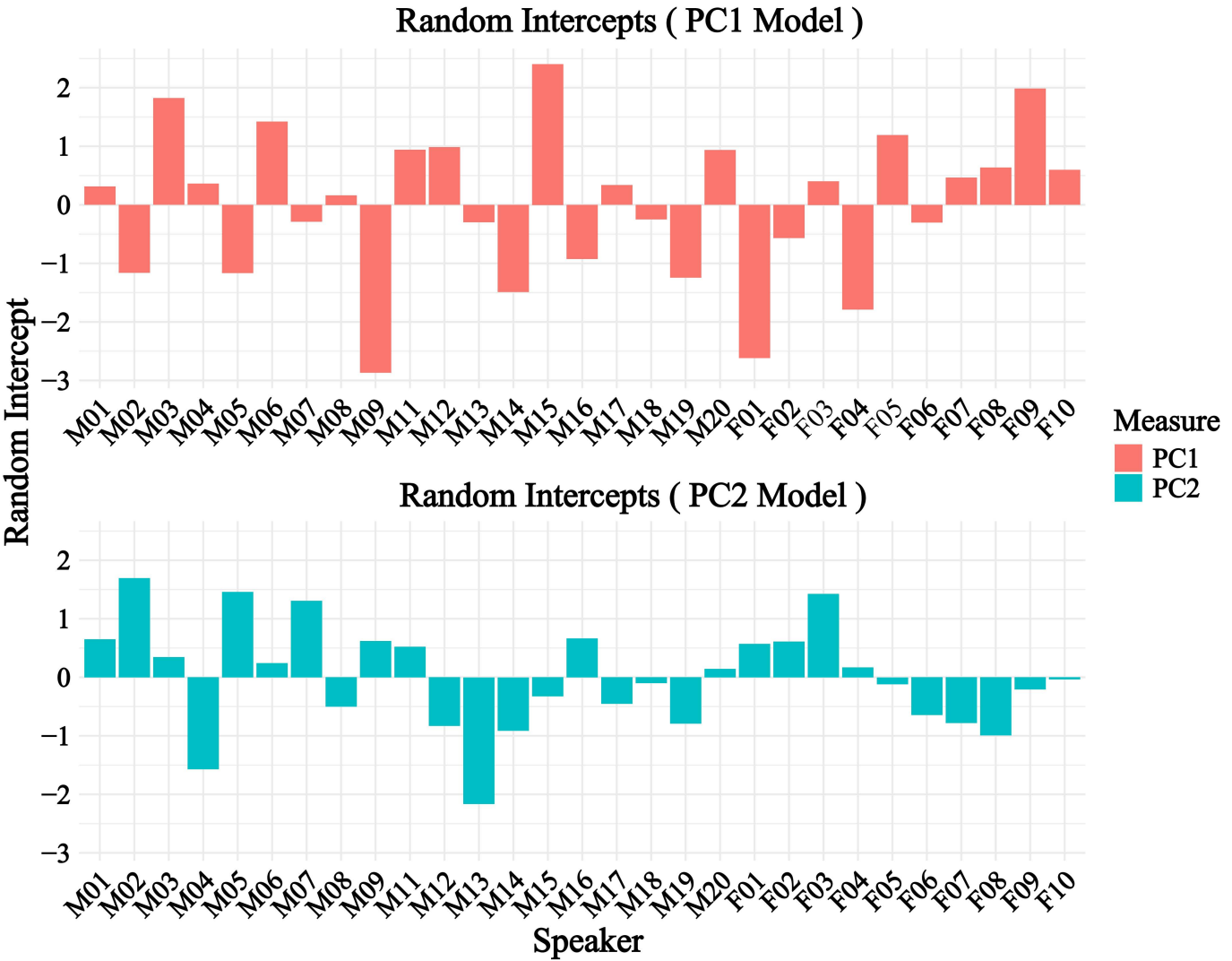

H1*-H2*
H1*-A1*
H1*-A2*
H1*-A3*
CPP
HNR05
HNR15
HNR25
HNR35
CQ
DECPA
Duration
dB
(unspecified units)
ms
T11 T22 T33 T44 T55
Tone
Gender
F
M

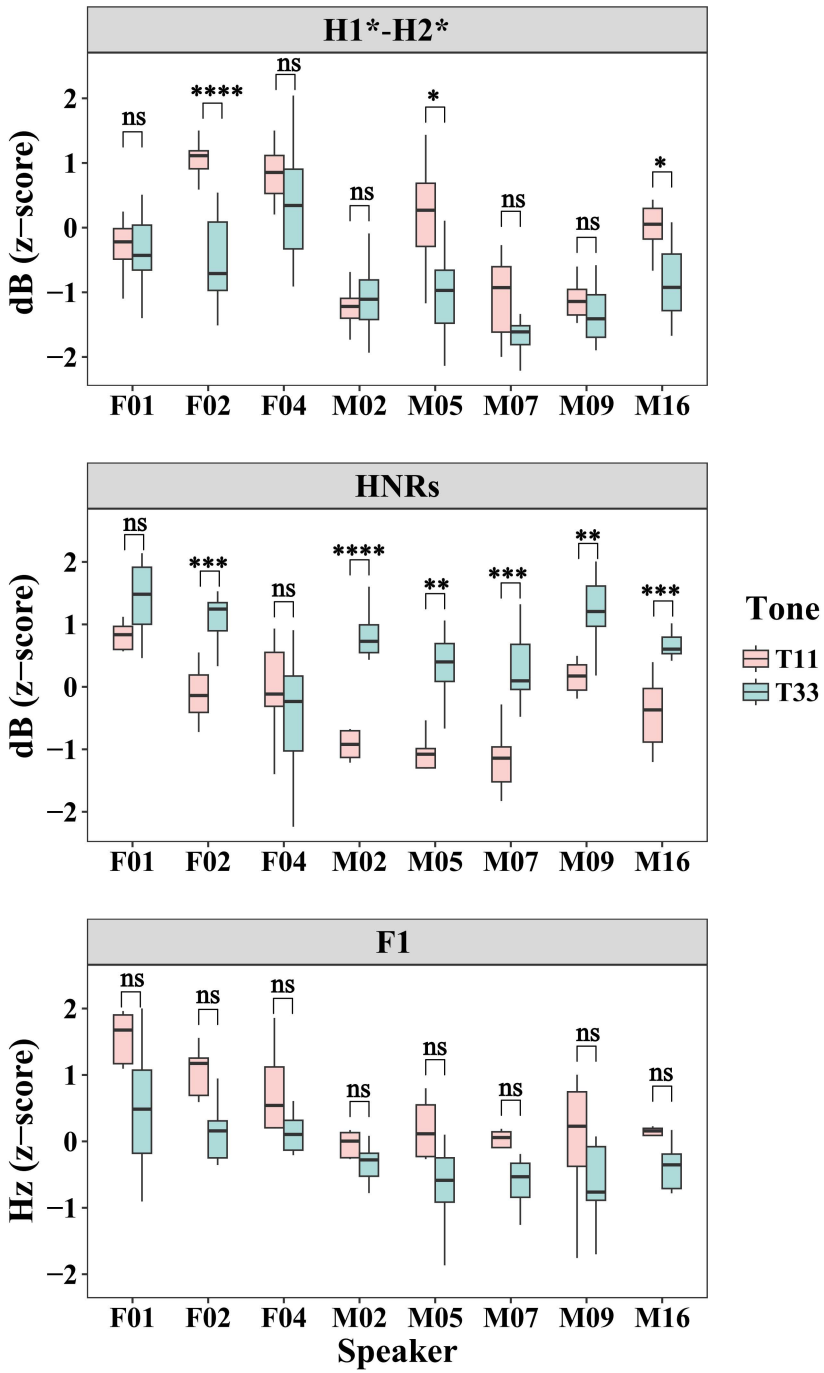

H1*-H2*
H1*-A1*
H1*-A2*
H1*-A3*
CPP
HNR05
HNR15
HNR25
HNR35
CQ
DECPA
dB
(unspecified units)
01
02
03
Gender
F
M

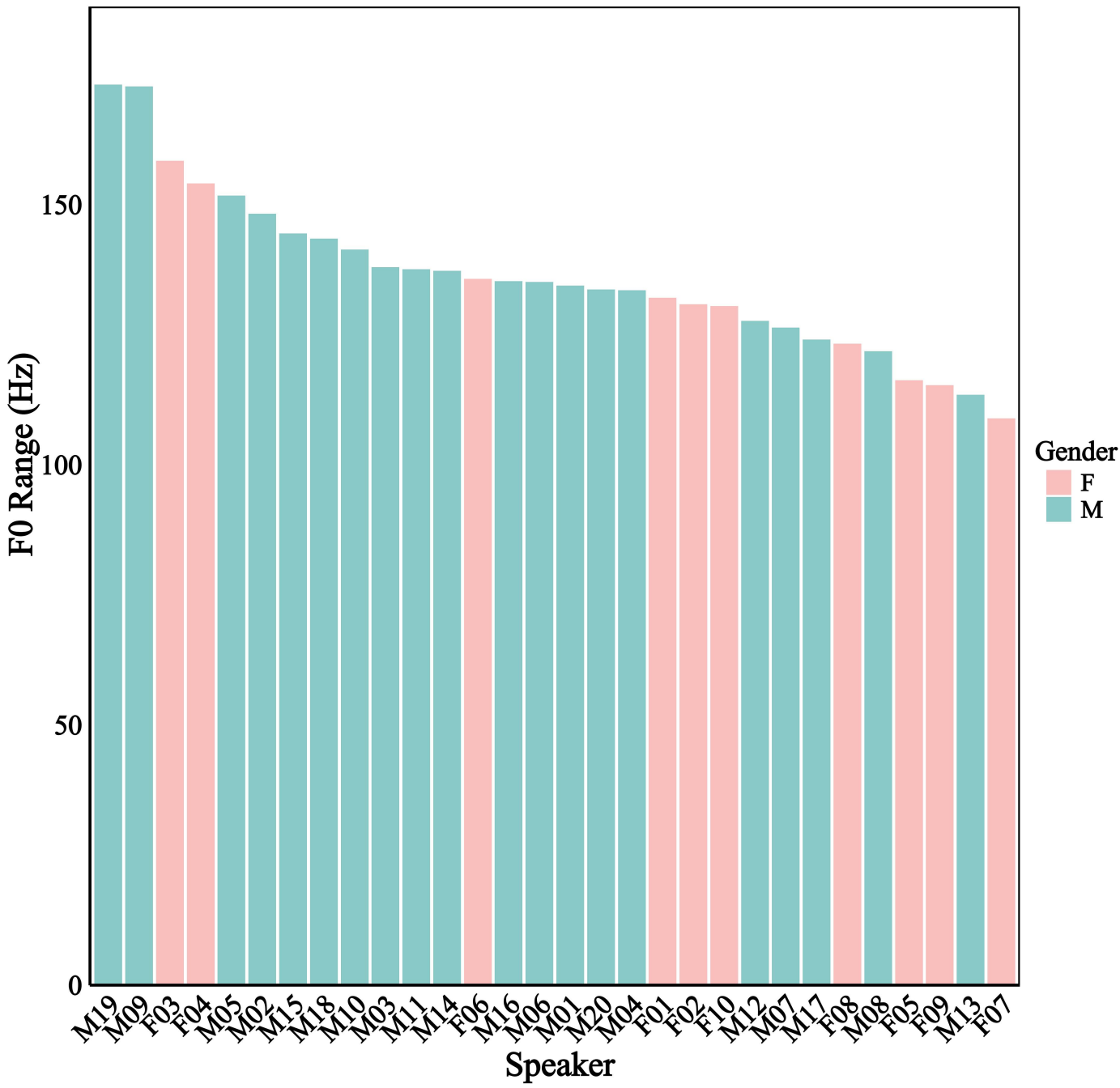

1.0
0.5
0.0
−0.5
−1.0
Dim2 (14.9%)
−1.0
−0.5
0.0
0.5
1.0
Dim1 (46.1%)
CQ
DECPA
CPP
HNR05
HNR15
HNR25
HNR35
H1*-A3*
H1*-A2*
H1*-A1*
H1*-H2*
Contrition
12.5
10.0
7.5
5.0
2.5


2.5
0.0
−2.5
−5.0
PC2
−4
0
4
PC1
Tone
T11
T22
T33
T44
T55


5.0
2.5
0.0
−2.5
−5.0
PC1 (predicted value)
20
30
40
50
60
Age (years)
Gender
F
M




H1*-H2*
dB (z−score)
ns
****
ns
*
ns
ns
ns
*
F01
F02
F04
M02
M05
M07
M09
M16
HNRs
dB (z−score)
ns
***
ns
****
**
***
**
***
Tone
T11
T33
F1
Hz (z−score)
ns
ns
ns
ns
ns
ns
ns
ns
Speaker


F0 Range (Hz)
150
100
50
0
M19
M09
F03
F04
M05
M02
M15
M18
M10
M03
M11
M14
F06
M16
M06
M01
M20
M04
F01
F02
F10
M12
M07
M17
F08
M08
F05
F09
M13
F07
Speaker
Gender
F
M