JASA ARTICLE

Harsh voice and its interaction with vowel quality in Fuzhou Min Chinese

Changhe Chen and Jonathan Havenhill[a)]
Department of Linguistics, The University of Hong Kong, Pokfulam, Hong Kong

ABSTRACT:
Fuzhou Chinese (Fuzhounese), a variety of Eastern Min, is known for its tone-vowel interaction. The tones /21, 241, 24/ are associated with lower, diphthongal vowels and non-modal phonation, while tones /44, 51, 32, 5/ are associated with higher, monophthongal vowels and modal phonation. This study tests the hypothesis, based on the laryngeal articulator model (Esling *et al.*, 2019), that epilaryngeal constriction responsible for non-modal phonation mediates vowel quality alternation due to articulatory synergy with lingual retraction. The most common phonation type observed among 18 speakers is harsh voice (with occasional aryepiglottic trilling). Creaky and modal voice, and less often ventricular or whispery voice, are also observed. Spectral tilt and noise measures reveal that noise (cepstral peak prominence and harmonic-to-noise ratio) is most reliable in distinguishing modal from non-modal tones. The only spectral tilt measure that does so is H2–H4, while H1–H2 differentiates subtypes of constricted phonation. The relationship between phonation and vowel quality is also examined. Noisier phonation is associated with higher F1 for all vowels, lower F2 for [ei], and higher F2 for [ou], in accordance with cross-dialectal variation (Chen, 1998), supporting the hypothesis that tone-vowel interaction is mediated by epilaryngeally constricted phonation. https://doi.org/10.1121/10.0036256

(Received 23 May 2024; revised 28 February 2025; accepted 6 March 2025; published online 8 April 2025)

[Editor: Jianjing Kuang] Pages: 2582–2602

I. INTRODUCTION

Tone contrasts are often realized by multiple acoustic cues. For example, the low dipping tone of Mandarin and the low falling tone of Cantonese are typically produced with creaky voice (e.g., Kuang, 2017; Yu and Lam, 2014), while the low register tones of Wu dialects are produced with diverse voice qualities (e.g., Gao and Hallé, 2017; Gao *et al.*, 2020; Ge *et al.*, 2023; Rose, 2015; Tian and Kuang, 2021; Xu and Mok, 2021). In Sinitic, the use of a three-way combination of pitch, voice quality, and vowel quality for tonal contrast is exemplified by some varieties of Eastern Min. In Fuzhou Chinese (Fuzhounese), which is considered the representative variety, the tones /44, 51, 32, 5/ are produced with modal voice (Donohue, 2017; Hisao, 2010) and with higher, more monophthongal vowel qualities. The tones /21, 241, 24/, on the other hand, are produced with "creaky voice" (Donohue, 2017) or "glottalization" (Hisao, 2010) and are accompanied by a lower and more diphthongal vowel quality. This sort of association between laryngeal and supralaryngeal articulation indicates that vowel and tone production are not strictly separated, which can be understood through the laryngeal articulator model (LAM) (Esling *et al.*, 2019). The LAM proposes that the production of constricted voice qualities is accompanied by the adjustment of tongue position, which in turn may lead to a change in vowel quality. This model successfully accounts for similar patterns in Nuosu Yi, a Lolo-Burmese language (Edmondson *et al.*, 2017), and Somali, a Cushitic language (Edmondson and Esling, 2006). The aims of the present study are to provide new insight into tone-vowel interaction in Fuzhounese (Donohue, 2017; Hisao, 2010; Jiang-King, 1996; Maddieson, 1976; Yip, 1980) and to enhance understanding of the relationship between tone, voice quality, and vowel quality more generally.

A. Voice quality and vowel quality

Voice quality, used here to refer to phonatory/laryngeal quality (compare Esling *et al.*, 2019; Laver, 1980), is known to interact with vowel quality. Cross-linguistic comparisons reveal that vowels with lax or breathy voice tend to have a higher vowel quality (more close, lower F1) than vowels with modal voicing, as in the Nilotic language Dinka (Denning, 1989), the Austroasiatic languages Eastern Cham and Chrau (Brunelle, 2005; Tạ *et al.*, 2022), and the Tibeto-Burman language Southern Yi (Kuang and Cui, 2018), among others. Similarly, the register system of Javanese (Austronesian) distinguishes onset stop consonants not only by their voicing, but also by differences in pitch, voice quality, and vowel quality (Fagan, 1988; Thurgood, 2004). These patterns can be explained, in part, by the perceived auditory quality of breathy-voiced vowels, wherein increased spectral tilt and stronger H1 interact with F1 to contribute to the perception of vowels as having a more close quality (e.g., Lotto *et al.*, 1997).

[a)]Email: jhavenhill@hku.hk

https://doi.org/10.1121/10.0036256

On the other hand, the association of creaky voice (and other types of constricted phonation) with lower vowel height is better explained by articulatory factors under the LAM (Esling, 2005; Esling *et al*., 2019). In contrast to models in which phonation is characterized by openness of the glottis (e.g., Ladefoged, 1971, reviewed by Gordon and Ladefoged, 2001), the LAM emphasizes the interdependence of the tongue, jaw, and larynx in modifying the shape of the entire lower vocal tract. This relationship accounts for findings (documented since Hollien, 1972) that creaky voice not only involves more abrupt adduction of the vocal folds and a longer closure phase, but also additional epilaryngeal constrictions, including incursion of the ventricular folds, aryepiglotto-epiglottal constriction, or epiglotto-pharyngeal constriction. Because the tongue and the larynx are biomechanically linked by the epilarynx, laryngeal constriction promotes lowering and retraction of the tongue body, thereby yielding a more open vowel quality.

This articulatory connection can explain a number of cross-linguistic sound patterns in which lingual and laryngeal features are observed to interact. For instance, Edmondson and Esling (2006) compare the non-constricted Somali vowels [i iː e eː æː ɐ o oː u uː] with their constricted counterparts [ɪ ɪː ɛ ɛː ɑː a ɔ ɔː ʊ ʊː], examining vowel formants and laryngeal articulation. Accompanied by harsh or creaky voice, constricted vowels are shown to be produced with ventricular fold incursion, increased constriction of the pharynx and of the arytenoid-epiglottal aperture, as well as lowering and backing of the tongue and epiglottis. Along with laryngeal raising, which facilitates constriction, these actions conspire to shorten the pharynx and thus have ramifications for acoustic vowel quality. As such, the constricted vowels have a higher F1, while F2 may be higher or lower depending on which corner of the vowel space the vowel is situated in. Vowel quality differences between the two sets of Somali vowels are similar to those described for Eastern Min. However, voice quality in Eastern Min and its relationship to vowel quality have not been systematically investigated.

B. The sound system of Fuzhounese

In this study, we examine tone-vowel interaction in Fuzhounese, an Eastern Min language predominantly spoken in Fuzhou (also known as Foochow or Hokchiu), the capital city of China's Fujian Province. Fuzhounese is also commonly spoken by immigrant communities in the Sai Wan and Tsuen Wan districts of Hong Kong; Chinatown in Manhattan, New York City; and Sibu, Malaysia, among other locations.

The consonant inventory of Fuzhounese is given in Table I. All consonants can appear in onset position, while codas are limited to /ŋ/ and /ʔ/. Symbols given in parentheses indicate allophones that emerge due to lenition. Specifically, for syllables in the final position of a disyllabic word, the initial consonant is often deleted, lenited, or assimilated. For example, /kou^{21}/ + /søy^{241}/ surfaces as [ku^{44} ɾøy^{241}] "story," with /s/ leniting to [ɾ], while /tɔy^{21}/ + /ts^{h}iu^{32}/ surfaces as [tøy^{53} iu^{32}] "to help," with deletion of /ts^{h}/. These processes occur alongside right-dominant tone sandhi, which determines the tone and the vowel of the penultimate syllable.

TABLE I. Consonant inventory of Fuzhounese. Allophones appear in parentheses.

	Bilabial	Alveolar	Palatal	Velar	Glottal
Stop	p p^{h}	t t^{h}		k k^{h}	ʔ
Approximant/tap	(β̞)	l, (ɾ)	(j)		
Nasal	m	n		ŋ	
Fricative/affricate		s ts ts^{h}			h

Fuzhounese has seven citation tones, as given in Table II, that occur freely in the final position of a tone sandhi domain. The tonal inventory includes two high level tones—checked (/5/) and unchecked (/44/); three falling tones—high (/51/), mid (/32/), and low (/21/); a low rising tone (/24/); and a low rise-fall (/241/). Tones can be organized into two categories (or registers), referred to here as Set A and Set B, according to their participation in tone-vowel interaction and presence of non-modal phonation, discussed in more detail below. Average pitch contours for all seven tones are given in Fig. 1. The low rising tone /24/ and the high-level tone /5/ occur only in short syllables checked by a glottal stop coda, and can therefore be distinguished from /241/ and /44/, respectively, by differences in both duration and F0.[1] These two cues have been the exclusive focus of prior tonal acoustic studies of Fuzhounese (Donohue, 2013; Peng, 2011) and Fuqing Chinese (Lam, 2014).

C. Tone-vowel interaction in Eastern Min

The tone sandhi systems of Eastern Min have garnered attention for their vowel quality alternations that accompany tone change (Chan, 1985; Wright, 1983; Yip, 1980). In disyllabic words, the tone of the final syllable retains its citation form, while the penultimate syllable changes according to tonal context. This results in widespread neutralization such that non-final Set A and Set B tones collapse either into Set A tones or non-citation tones ostensibly belonging to neither register (Chen, 1998; Feng, 1998; Liang, 1986; Peng, 2011).

TABLE II. Tone and vowel inventory of Fuzhounese in final/isolated syllables. MC indicates correspondences to Middle Chinese tone categories based on historical voicing (+, yang) or voicelessness (–, yin) of the onset.

	Tone	MC	Phonation type	Vowel qualities
Set A	˦ 44	—	modal	i y u
	˥ 5	+		e ø o
	˥˩ 51	+		a
	˧˨ 32	—		
Set B	˨˩ 21	—	constricted	ei øy ou
	˨˦˩ 241	+		a ɔ ɔ
	˨˦ 24	—		a

https://doi.org/10.1121/10.0036256

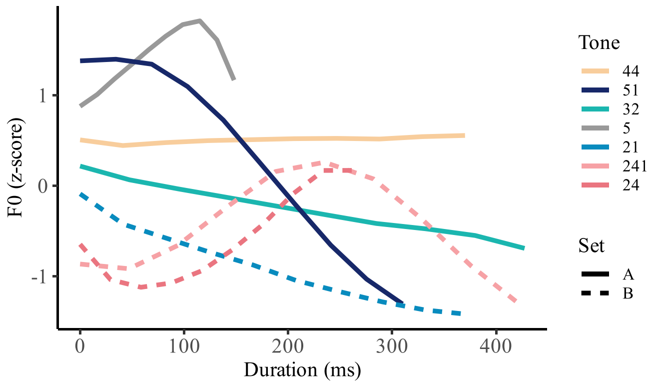

FIG. 1. z-score normalized pitch contours for the seven tones of Fuzhounese, averaged for all speakers in the present study. Measurements were sampled from syllables containing the vowel /o/ ([o^{44}], [oʔ^{5}], [o^{51}], [o^{32}], [ɔ^{21}], [ɔ^{241}], [ɔʔ^{24}]) at 10% intervals and standardized according to the average duration for each tone.

In many varieties of Eastern Min, syllables carrying a Set B tone also exhibit a lower, retracted, and/or diphthongized vowel quality. As shown in Table II, the Fuzhounese vowels /i y e ø o u/ are realized either as [i y e ø o u] or [ei øy a ɔ ɔ ou], depending on the tonal context. In syllables containing the Set A tones /44, 51, 32, 5/, the higher, monophthongal allophones appear. When accompanied by the Set B tones /21, 241, 24/, the high vowels /i, y, u/ are typically diphthongized to [ei, øy, ou], lowered /ø, o/ are both realized as [ɔ], and lowered /e/ is realized as [a]. The underlyingly low vowel /a/ does not alternate (Peng, 2011). Because vowel quality alternations co-occur with tone sandhi, vowels retain their canonical (raised) quality when they appear in non-final position. For example, Fuzhounese 味 "taste" is realized as [ei^{241}] in final position or in isolation, such as in 殕味 "musty smell" [p^{h}u^{44} ei^{241}], but is realized as [i^{53}] in 味素"MSG" [i^{53}ɾou^{21}] in non-final position. That is, when the rising-falling tone /241/ changes to high falling [53] as the result of tone sandhi, the vowel is realized as high [i] rather than as its lowered/diphthongized allophone [ei].

Based on interdialectal phonetic variability within Eastern Min, shown for /i/ in Table III, Chen (1998) proposes that the diachronic development of vowel lowering involves several stages, i.e., [i]→[ɪ]→[e]→[ei]→[æi]. This progression is reflected across present-day varieties of Eastern Min, in which the lowered allophones of /i/ may be realized as [i], [ɪ], [e], [ei], or [æi], with the vowel nucleus lowering to varying degrees. Chen's proposal is consistent with contemporary and historical descriptions of Eastern Min vowels throughout the past 400 years. A 17th-century description of Eastern Min phonology, the rhyme book *Qi Lin Bayin* 戚林八音 ("Book of Eight Sounds"), suggests that there was no tone-vowel interaction at that time. This is apparent from words categorized in that work as containing the same onset and rhyme (differing only by tone), which are now distinguished by both tone and vowel quality. For instance, 賣 "to sell" and 買 "to buy" are realized in present-day Fuzhounese as [ma^{241}] and [me^{32}], yet were classified as having an identical onset and rhyme but different tones (i.e., for expository purpose, [me^{241}] and [me^{32}]).

TABLE III. Allophones of /i/ in 19 varieties of Eastern Min, as reported by Chen (1998). Lower allophones occur exclusively with particular tone categories, which vary on a variety-specific basis.

Allophone	Eastern Min varieties
[i]	Gutian, Luoyuan, Songkou, Bandong, Yanping (no tone-vowel interaction)
[ɪ]	Tandong
[e]	Fuqing, Longtian, Pingtan
[ei]	Fuzhou,[a] Shanyang, Tangkou, Daixi, Lianjiang, Xiao'ao, Yongtai, Changle, Minhou
[æi]	Minqing

[a]Focus of the current study.

Maclay and Baldwin (1870) report that the Fuzhounese vowel "á" (transcribed as /e/ in the present study) was consistently realized as [e] and was not reported to have a lowered allophone at that time. They do identify an alternation between higher/monophthongal and lower/diphthongal allophones for the high vowels /i y u/, suggesting that lowering of the high vowels preceded lowering of the mid vowels, a point to be revisited in the discussion. Tao (1930) proposed that the vowel /e/ exhibited two allophones, [ɛ] and [æ], with the latter being transcribed as low central [a] in more recent work (Chen, 1998; Feng, 1998; Liang, 1986). Synchronic interspeaker variability in vowel height has been observed in the Fuqing variety—Lam and Hong (2016) report that the lax vowel [ɪ] (transcribed as [e] by Chen, 1998) is further lowered and diphthongized to [ei] for some women. We observe similar inter- and intra-speaker variation with respect to the height of the Fuzhounese lax vowels [ei øy ou], which may also be realized as [ɛi œy ɔu].

D. Accounting for tone-vowel interaction in Eastern Min

Yip (2002) concludes that existing phonological accounts of Fuzhounese tone-vowel interaction have been "murky" and "controversial." She (and most others) consider the interaction between tone and vowel quality to be direct; that is, the tone itself is the trigger for vowel quality alternation. An alternative analysis, in which vowel quality alternation is the trigger for tone sandhi, is incompatible with the fact that the Set A tones also undergo tone sandhi but do not exhibit vowel quality alternation. To formalize this distinction, most analyses classify the tones into two registers defined by pitch height—Set A tones comprise the upper register and Set B tones the lower. This approach relies on the observation that the Set B tones /21, 241, 24/ begin with a relatively low pitch target, while the Set A tones /44, 51, 32, 5/ are higher. Some early accounts characterized tone-vowel alternation as vowel raising in non-final position (Mohr, 1971; Wang, 1972; Yip, 1980), possibly due to phonologization of intrinsic F0, a phenomenon in which high vowels tend to have a higher F0 than low vowels (Chen *et al.*, 2021; Whalen and Levitt, 1995). These accounts do not capture all reported vowel alternations (Jiang-King, 1996; Maddieson, 1976; Wright, 1983), however, nor is vowel raising supported by historical or cross-dialectal evidence already introduced. Later accounts propose that the low register Set B tones induce vowel lowering in prosodically prominent final position, due either to low pitch (Chan, 1985; Wright, 1983) or greater syllable weight (Jiang-King, 1996).

For most varieties of Chinese, which distinguish tones predominantly (or exclusively) by pitch, the register feature [±upper] serves to bifurcate tonal inventories on the basis of overall pitch height (Yip, 1980). For instance, the Cantonese low register tone /23/ (Tone 5) has a lower F0 but similar contour as compared to its high register counterpart /25/ (Tone 2) (Bauer and Benedict, 1997). While such a feature can also capture the synchronic patterning of Fuzhounese, it is not clear, on the whole, that Eastern Min register is straightforwardly associated with pitch. The Fuzhounese Set B tones all begin with a (relatively) low target—similarity of "high" register /32/ and "low" register /21/ notwithstanding—but the same is not necessarily true of other varieties with similar alternations. In Fuqing Chinese, for instance, the vowels /i y e ø ɐ o u/ are realized as [ɪ ʏ ɛ œ a ɔ ʊ] when accompanied by the tones /21, 41, 22/ (Lam, 2014; Lam and Hong, 2016). These historically correspond to Fuzhounese /21, 241, 24/, respectively, but present-day /41/ and /241/ clearly differ in whether the initial target is high or low. Chen (1998) indicates that vowel quality alternations in some varieties may be associated with different tone categories altogether: In Yongtai and Minqing, tones corresponding to Fuzhounese /21/ and /241/ (but not /24/) exhibit lower vowels, while in Shanyang, Tangkou, and Daixi, the tones associated with vowel alternation correspond to Fuzhounese /51/ and /241/. Detailed transcriptions or acoustic data for these varieties are not available, so evidence for an association of pitch height and vowel quality is inconclusive.

More broadly, the term "register" is used to describe systems in which prosodic categories differ along several correlated dimensions, including pitch, duration, vowel quality, voice quality, and/or consonant voicing. In Shanghainese and other varieties of Wu Chinese, the upper and lower registers are redundantly specified: High register tones have higher pitch, co-occur with phonologically voiceless onsets, and use modal voicing; low register tones have lower pitch, occur with voiced onsets, and use breathy, whispery, or creaky non-modal phonation (Gao and Hallé, 2017; Gao *et al.*, 2020; Rose, 2015; Tian and Kuang, 2021). In a phylogenetically diverse set of Southeast Asian languages (reviewed by Brunelle and Kirby, 2016), register contrast often also involves vowel quality alternation. In these languages, high registers tend to be realized by higher pitch, tense or modal voicing, vowels with a more open or monophthongal quality, and (historically) voiceless onsets. By contrast, low registers exhibit lower pitch, breathy voicing, vowels with a more close or diphthongized quality, and (historically) voiced onsets.

The register system proposed for Fuzhounese differs from these examples in several respects. For one, the proposed relationship between vowel height and pitch height opposes that observed in most Southeast Asian register systems, as the ostensibly low register Set B tones co-occur with more open (rather than more close) vowel qualities. Moreover, the Fuzhounese Set B tones are not known to correspond either to an extant or historical consonant voicing contrast, which is perhaps the most common source of tono- and registro-genesis cross-linguistically (Hombert *et al.*, 1979). Tones in most Chinese languages can thus be categorized according to onset voicing categories reconstructed for Middle Chinese: *yang* tones appear in syllables with historically voiced onsets and *yin* tones in those with historically voiceless onsets. In Wu, where onset voicing is preserved in some contexts (Chen and Gussenhoven, 2015), these categories directly correspond to the lower and upper registers, respectively. Contrastive voicing is no longer present in Cantonese, but the historical *yin* and *yang* categories neatly divide [±upper] tone pairs, such as /25/ vs /23/, that differ in overall pitch height. Fuzhounese tones are traditionally described in a similar fashion, as illustrated in Table II, yet this dichotomy exhibits no apparent relationship to the present-day tone classes. The Set B tones /21, 24/ occur in syllables with historically voiceless onsets while the Set B tone /241/ appears in syllables with historically voiced onsets. The Set A tones are similarly mixed and it is notable that Set A /32/ and Set B /21/ not only have similar pitch heights, but also both occur in words with historically voiceless onsets.

Register contrast nevertheless encompasses a diverse range of vowel, tonal, and/or phonatory qualities, which Esling *et al.* (2019) argue can all be classified by the presence or absence of epilaryngeal constriction and (anti-)synergy of associated lingual, laryngeal, and epilaryngeal states. As already indicated, Fuzhounese vowel alternations resemble a type of register contrast exemplified by Nuosu Yi and Somali, in which lower and more retracted vowels correlate with constricted phonation. The possible involvement of non-modal phonation in Fuzhounese tone-vowel interaction has been mostly overlooked until recently, however. The use of "creaky" or "breathy" phonation with the Set B tones was observed by Chan (1985), but she notes that both the phonetic data and theoretical frameworks available at the time were insufficient to properly describe the relationship. More recently, Lam (2014) has reported that the Set B-equivalent tones in Fuqing Chinese are produced with "creaky" voice, such that /41/ is realized with lowered (monophthongal) vowels and creaky phonation, which distinguishes it from /51/ (realized with high vowels and modal voicing). Donohue (2017) proposes that Fuzhounese tone-vowel interaction results from the enhancement of tonal register contrast with "creaky," "creaky/breathy," or "glottalized breathy" phonation (Donohue, 2011, 2013). Following earlier work, she considers Set B to comprise a class of low register ([–upper]) tones, which are realized with non-modal phonation and lower vowel qualities.

She argues that the presence of non-modal phonation serves to phonetically enhance contrast with [+upper] Set A tones. The phonetic mechanisms behind this relationship are not elaborated, however, most importantly, the basis for vowel quality alternation. Similarly, Hisao (2010) proposes that "glottalization" serves to enhance the distinction between low and high tones, but further hypothesizes that glottalization involves lowering of the jaw, which contributes to a lower tongue position.

Because none of these studies reports quantitative acoustic measures of voice quality, the phonetic characteristics of non-modal phonation associated with /21, 241, 24/ remain opaque. As previously noted, creaky or glottalized voicing belong to a broader category of phonation types involving epilaryngeal constriction (Esling *et al.*, 2019); the presence of noise in addition to creakiness, both correlates of epilaryngeal constriction, is suggested by Donohue's (2011) use of the term "glottalized breathy." This study therefore aims to characterize the specific non-modal phonation types used with /21, 241, 24/ and to determine which acoustic voice quality cues best differentiate the Set A and Set B tones. We hypothesize that the relationship of tone, vowel quality alternation, and non-modal phonation in Fuzhounese is driven by epilaryngeal constriction, which has articulatory consequences for both phonatory quality and vowel quality, as postulated by the LAM (Esling *et al.*, 2019). Given that the quality of the Set B vowels in Eastern Min is reported to vary synchronically (Lam and Hong, 2016; Peng, 2011), we predict that a greater degree of laryngeal constriction—aryepiglottic tightening of the epilaryngeal tube beneath the epiglottis—will contribute not only to increased frication, trilling, or occlusion, but will also promote more extreme lowering and retraction of the tongue, in line with the historical trend of lowering.

E. Research questions and hypotheses

Through the following acoustic analyses, we aim at investigating, first, the phonetic features of the phonation types associated with the Set B tones /21, 241, 24/, and second, the role of non-modal phonation in Fuzhounese tone-vowel interaction. Specifically, we are interested in the phonetic characteristics of Fuzhounese speakers' non-modal phonation and the acoustic correlates that can distinguish the Set A vs Set B tones. Based on the LAM, we hypothesize that constricted non-modal phonation, which has previously been described as "creaky" or "glottalized," mediates the tone-vowel interaction. If this hypothesis holds, we expect to observe the tones /21, 241, 24/ not only being produced with more constricted voice qualities, but also an inverse correlation of vowel height (as measured by F1) with harmonic-to-noise ratio (HNR). This follows from the assertion that a higher degree of epilaryngeal constriction should favor a lower tongue position. Regarding the relationship of non-modal phonation to F2, greater lingual lowering is expected to result in vowel-specific effects, according in part to their backness, as has been shown for constricted vs non-constricted vowels in Somali (Edmondson and Esling, 2006).

The analysis is organized as follows. Section III A describes the phonation types observed for the Set B tones, which range from modal to harsh, and their frequency of occurrence on a speaker-specific and tone-specific basis. Section III B examines the acoustic correlates of the various phonation types observed for Set B tones, as well as the ability of spectral tilt and noise measures to distinguish the Set A and Set B tones. Section III C investigates the relationship between vowel quality and constricted phonation, while Sec. IV discusses the implications of these findings for our understanding of tone-vowel interaction in Eastern Min and the relationship between voice and vowel quality.

II. METHODS

A. Participants

Eighteen native speakers of Fuzhounese (7 men, 11 women) were recorded. Participant birth years are shown in Table IV, with all but one participant having been born and raised in Fuzhou at least through age 18. While all participants also speak Mandarin, they report predominant use of Fuzhounese in their households and with friends. FZ07M was raised in a multigenerational Fuzhounese-speaking household in Jian'ou, Fujian Province; he also speaks Jian'ou Chinese, a mutually unintelligible variety of Northern Min which lacks tone-vowel interaction. FZ09F and FZ11F relocated to Hong Kong after age 18 and also speak Cantonese. Attempts to recruit additional male speakers were unsuccessful; two respondents did not follow up with us, eight were ultimately uninterested in participating, and one participant found the elicitation task to be tedious and did not complete the session. No participant reported any speech or hearing disorders.

B. Materials and procedure

Participants recited a list of 153 monosyllabic words representing the full inventory of (historically) monophthongal Fuzhounese vowels, /i y u e ø o a/, as well as phonemic diphthongs. Both the canonical Set A and lowered/retracted/diphthongized Set B allophones are represented, including 43 (C)V or (C)Vʔ words containing Set A tones and 61 containing Set B tones. Forty-nine filler items were included for the purposes of another study. Speakers were asked to repeat

TABLE IV. Birth year and gender of participants in this study.

Decade	Speakers
1940s	FZ03M (1945)
1950s	FZ02M (1955), FZ03F (1950), FZ04M (1957), FZ06F (1950), FZ06M (1959), FZ07M (1957), FZ08F (1956), FZ11F (1954)
1960s	FZ01F (1969), FZ01M (1960), FZ02F (1962), FZ04F (1963), FZ05F (1960), FZ10F (1960)
1970s	FZ09F (1974)
1980s	FZ07F (1987)
2000s	FZ05M (2001)

the words in a randomized order, with each word spoken in isolation to elicit the tones' citation form and to avoid consonantal alternations in the target syllable. The entire word list was repeated three times. Recordings were carried out in a quiet room or sound-attenuated booth with speakers seated at a desk or table. Recordings were captured using the internal microphone of a Zoom H5 recorder, which was chosen for its flat frequency response. The recorder was mounted to a desktop stand at a height approximately level with the speaker's mouth at a distance of around 20 cm.

Given that the focus of this study is on the tone-vowel alternation that occurs exclusively with Set B tones, all 61 Set B words are included in the analysis, listed in Tables X and XI (see the Appendix). Of the Set A words, 10 words containing the phonemic mid and low vowels /e o a/, given in Table X, were analyzed. These words were chosen for comparison with the Set B items because their canonical vowel quality is comparable to the nuclei of their Set B counterparts [ei ou a]. The remaining items read by participants are therefore not directly considered at present. In particular, the vowel allophones [i y u] are excluded because phonetically high vowels can only occur with the Set A tones and are less straightforwardly compared to the Set B vowels, all of which are phonetically mid or low. Moreover, some speakers produce [i u] with frication, which can be expected to bias the measurement of noise generated in the lower vocal tract. Similar frication of high vowels can be observed in other Sinitic varieties, including some dialects of Hui, Jin, Lower Yangtze Mandarin, Northwestern Mandarin, and Wu (for reviews, see Faytak, 2014; Shi, 1998).

C. Data analysis

Spectral tilt and noise measures were obtained using VoiceSauce (Shue *et al.*, 2011). Spectral tilt measures included H1*–H2*, H2*–H4*, H1*–A1*, H1*–A2*, H1*–A3*, H4*–H2k*, and H2k*–H5k*. Here, H1, H2, H4, H2k, and H5k refer to the amplitudes of the first, second, and fourth harmonics and to the harmonics near 2000 and 5000 Hz, respectively. A1, A2, and A3 are the amplitudes of the harmonics with the highest amplitude near the first, second, and third formants, respectively. Asterisks indicate that harmonic amplitudes were corrected for the influence of upper (oral) vocal tract resonances, allowing for cross-vowel comparisons.

Several of these measures have been demonstrated to relate to the manner of vocal fold vibration (for reviews, see Chai and Garellek, 2022; Tian and Kuang, 2021), which has traditionally been considered the primary determinant of phonation type (Ladefoged, 1971). Lower H1–H2, for example, is correlated with a lower proportion of glottal opening (opening quotient) in each glottal cycle (Holmberg *et al.*, 1995) and increased thickness of the medial vocal fold (Zhang, 2016). Lower H1*–A1* is correlated with a smaller glottal opening at the arytenoids (Hanson *et al.*, 2001), while lower H1*–A2* and H1*–A3* are correlated with more abrupt adduction of the vocal folds (Hanson *et al.*, 2001; Stevens, 1977). Noise measures included cepstral peak prominence (CPP) as well as HNR in four frequency bands (HNR05: 0–500 Hz; HNR15: 0–1500 Hz; HNR25: 0–2500 Hz; and HNR35: 0–3500 Hz).

The ability of each of these acoustic measures to distinguish the Set B vs Set A tones is examined in Sec. III B: "Acoustic correlates of non-modal phonation." Visual inspection of the Set B tones /21, 241, 24/ reveals that non-modal phonation is typically associated with the rhyme onset, while the latter portion is generally produced with modal voicing. Therefore, voice quality and F0 measures were averaged over the first one-third interval of the target Set A and Set B tokens. A linear mixed effects regression model was constructed for each acoustic measure using the lme4 package (Bates *et al.*, 2015) in R (R Core Team, 2024) to test whether that measure can distinguish the two types of tones. The model formula is as follows: `measure ~ tone + F0 + (1 + tone | speaker) + (1 | word)`. Because the idiosyncrasy of each speaker is captured by the random intercept and slope, models were fit to the raw (unnormalized) data. ANOVA comparison of two models, with and without the fixed effect, determines whether the acoustic measure distinguishes the two types of tone. F0 is included in the model as a control for intrinsic covariation of F0 and spectral tilt/noise measures, which is observed even in the absence of phonologically specified differences in phonation (Kuang, 2017; Rose, 2020). Tokens were counted as mispronunciations and excluded from analysis in cases where the participant provided an unexpected character reading differing categorically from the target. For instance, attempts to elicit [øy^{21}] "to seep" often yielded the near-synonym [t^{h}au^{21}] given that [øy^{21}] lacks an orthographic character. In total, 480 Set A tokens and 474 Set B tokens were entered into the analysis.

The relationship between epilaryngeal constriction and vowel quality is examined for the Set B tones in Sec. III C: "Relationship between voice quality and vowel quality." CPP, HNR, and F0 measurements corresponding to the point of F1 maximum for [e ø o] (the nucleus of the Set B allophones [ei øy ou]) were selected for analysis. Values for both F1 and F2 were obtained at the same point. A linear mixed effects model was constructed to test whether the noise measures correspond to differences in F1 (vowel height) and F2 (vowel backness) for each vowel. The formula is as follows: `F1/F2 ~ measure + onset + F0 + measure:F0 + (1 + measure | speaker) + (1 | word)`. All measurements were z-score standardized; random speaker slopes for measure were excluded when issues with singular fit or failure to converge were encountered (noted in Sec. III). Correlation with F1 or F2 is determined by ANOVA comparison of two models fit with and without each noise measure. Mispronounced tokens are again excluded, providing a total of 2915 tokens (996 tokens of [ei], 705 tokens of [øy], and 1214 tokens of [ou]).

III. RESULTS

A. Phonation types for the Fuzhounese Set B tones

In order to determine the relative incidence of non-modal phonation, all tokens included in subsequent acoustic

https://doi.org/10.1121/10.0036256

analyses were categorized by phonation type, including both Set A and Set B tones. Voice quality characteristics were determined through visual inspection of the waveform and spectrogram, as well as auditory judgment. In addition to modal voice, multiple types of constricted voice quality were observed, including whispery voice, creaky voice, harsh voice (with or without aryepiglottic trilling), and ventricular voice. For the Set A tones, 95% ($N = 456$) were produced with modal voicing. Instances of non-modal phonation for Set A were sporadic; FZ04F and FZ03M were the only speakers who used harsh or creaky phonation for more than one (but fewer than five) Set A token. Because the primary concern of this study is the interaction of non-modal phonation with vowel quality alternation, the Set B tones involved in historical vowel lowering, /21, 241, 24/, are examined in detail.

The features which distinguish each phonation type are summarized in Table V, with corresponding examples given as Figs. 2–9. In all figures, the waveform excerpt corresponds to the interval indicated in the spectrogram. For visualization, pitch tracks were calculated in Praat (Boersma and Weenink, 2024) using the filtered autocorrelation method. Features examined included waveform periodicity, the presence or absence of turbulent noise, waveform damping, and the presence of secondary pulses. A typical modal voiced Set B vowel is shown in Fig. 2, in which the waveform is generally periodic, and the formant structure is clear and uninterrupted by noise.

TABLE V. Major features of the phonation types associated with the Set B tones /21, 241, 24/ in 2915 tokens.

Phonation type	Acoustic and auditory features	Examples
Modal	Waveform: periodic waveform with little or no noise Spectrogram: continuous clear formant structure Auditory features: neither tense nor lax	Fig. 2
Whispery	Waveform: waveform with noise and relatively low amplitude Spectrogram: continuous faint formant structure Auditory features: constricted and noisy	Fig. 3
Creaky	Waveform: periodic waveform with noise, high damping of the waveform Spectrogram: continuous clear formant structure Auditory features: constricted sound with distinct individual pulses	Fig. 4
Harsh	Waveform: periodic or aperiodic waveform with relatively strong noise Spectrogram: formant structure interrupted by noise Auditory features: constricted, strident or hoarse; possible aryepiglottic trilling	Figs. 5–8
Ventricular	Waveform: alternating strong and weak pulses Spectrogram: continuous, clear formant structure characterized by alternating strong and weak striations Auditory features: constricted sound with distinct individual pulses, perceived to have simultaneous high and low pitches or to have indeterminate/unclear pitch	Fig. 9

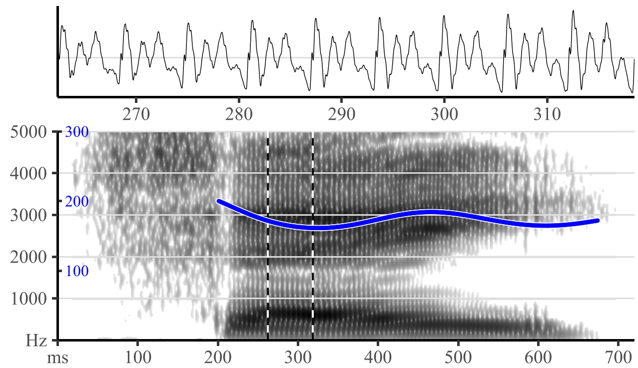

FIG. 2. Waveform excerpt, spectrogram, and pitch track for [sei^{241}] produced with modal voice by FZ07F.

Previous studies differ in their definition of "whispery voice" (for a review, see Tian and Kuang, 2021). We adopt the definition of Esling (2013), where whispery voice is produced with posterior opening of the glottis (similar to breathy voice), while at the same time epilaryngeal constriction is present. Thus, tokens with both a constricted and noisy auditory quality are categorized as "whispery voice" in the current study. In Fig. 3, the whispery voice waveform has high-frequency noise and a lower degree of periodicity (compared with modal voice). The spectral energy is stronger in the low-frequency region (below 900 Hz in this case) and the formant structure in the higher frequencies is faint. The constricted auditory quality may result from the combination of low pitch, low amplitude, and lowered vowel quality. The breathy auditory quality is characterized by stable airflow during voicing. Because whispery voice and breathy voice have the same manner of vocal fold vibration (Esling *et al.*, 2017), we expect that they will also measure the same, that is, higher values for measures of spectral tilt (e.g., H1–H2, H2–H4) and lower values for measures of spectral noise (e.g., HNR05, HNR15), reflecting the increased strength of turbulent noise (Garellek, 2019).

Creaky voice is also produced with epilaryngeal constriction; the vocal folds are shortened and thickened, and thus vibrate slowly (Esling, 2013). It can be seen from Fig. 4 that the formant structure is well defined, while the

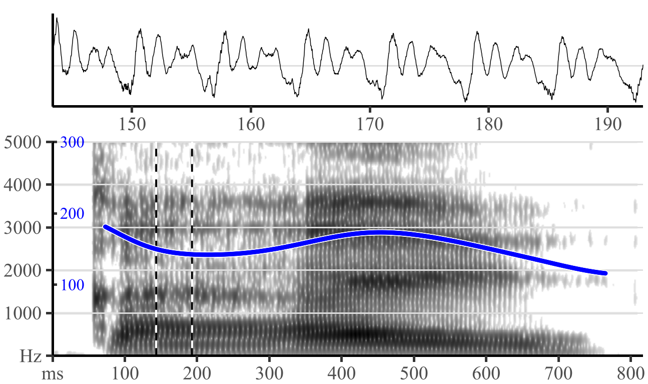

FIG. 3. Waveform excerpt, spectrogram, and pitch track for [køy^{241}] produced with whispery voice by FZ01F.

https://doi.org/10.1121/10.0036256

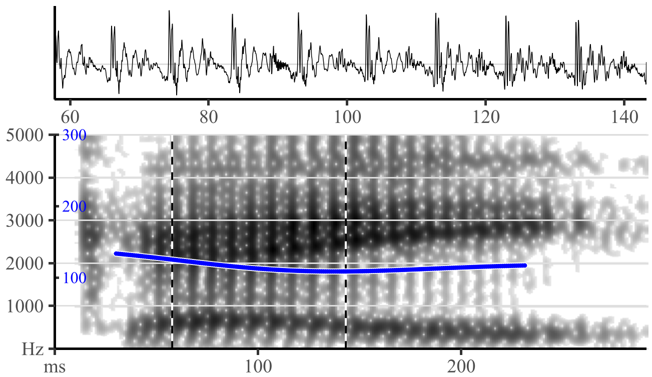

FIG. 4. Waveform excerpt, spectrogram, and pitch track for [keiʔ^{24}] produced with creaky voice by FZ10F.

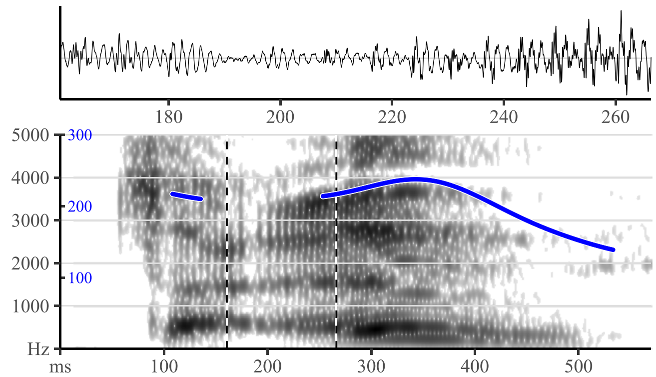

FIG. 6. Waveform excerpt, spectrogram, and pitch track for [tsøy^{241}] produced with harsh voice by FZ04F.

waveform shows a high degree of damping due to engagement of the ventricular folds (Moisik and Esling, 2014) and low rate of vibration (Keating *et al.*, 2015), in this case around 100 Hz. Individual glottal pulses are clearly audible. Creaky voice is associated with lower values for both spectral tilt and noise measures (Garellek, 2019), reflecting an overall decrease in harmonic energy and increase in noisiness due to irregular vocal fold vibration.

Harsh voice occurs when the airflow and epilaryngeal constriction of whispery or creaky voice are strengthened, leading to a noisier voice quality and/or the vibration of epilaryngeal constrictors (e.g., aryepiglottic folds) (Esling, 2013). Figures 5–8 show several different realizations of harsh voice in the sample. In Fig. 5, the formant structure is partially obscured by noise; F2 and higher formants are less clear. The waveform shows high variability in its amplitude, with several low-amplitude glottal cycles visible in the middle of the excerpted waveform. Figure 6 shows a token which sounds particularly constricted and exhibits a transient stop, where the amplitude is close to zero. Compared with Fig. 5, the formant structure in Fig. 6 is less obscured by noise. By contrast, Fig. 7 exhibits relatively less interruption of regular vocal fold vibration and has well-defined formant structure. However, each glottal pulse carries high-frequency noise. In comparison with whispery voice, this kind of harsh voice is characterized by a more strident and hoarse auditory quality. As seen in Fig. 8, harsh voice sometimes also occurs with aryepiglottic trilling, that is, vibration of the aryepiglottic folds. The most distinctive feature of this type of harsh voice is high-amplitude, low-frequency vibrations accompanied by hoarseness. This voicing type is also observed in some varieties of Wu and has been referred to as "growl" (Esling *et al.*, 2019; Rose, 2015). Based on these features, harsh voice is expected to have higher or lower spectral tilt measures, dependent on the manner of vocal fold vibration, and lower values for noise measures, with the latter corresponding to greater epilaryngeal constriction than for creaky or whispery voice.

Ventricular voice is defined by independent vibration of the ventricular folds simultaneous with true vocal fold vibration (Esling *et al.*, 2019). Because the two pairs of folds are aerodynamically entrained, the ventricular folds typically vibrate at half the rate of the vocal folds, yielding a waveform characterized by alternating strong and weak pulses, as seen in Fig. 9. As such, ventricular voicing has also been referred to as "period-doubled" (Bailly *et al.*, 2010; Huang, 2022) or "multiply pulsed" voice (Keating *et al.*, 2015). It may be auditorily perceived to have an indeterminate pitch (Keating *et al.*, 2015) or two distinct pitches, i.e., a low pitch superimposed by a second, higher pitch (Huang, 2022). In this example, the F0 during ventricular voicing is approximately 148 Hz, but additional, lower-amplitude

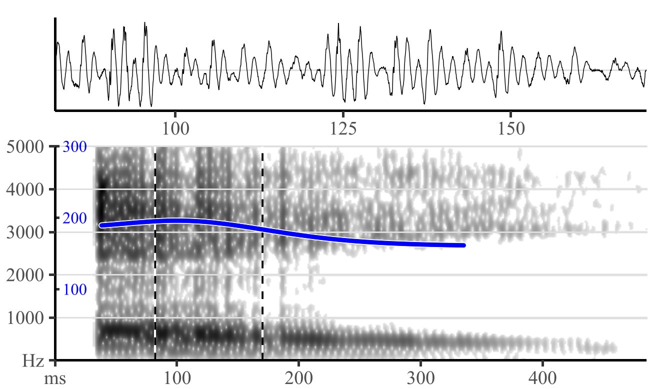

FIG. 5. Waveform excerpt, spectrogram, and pitch track for [ei^{21}] produced with harsh voice by FZ02F.

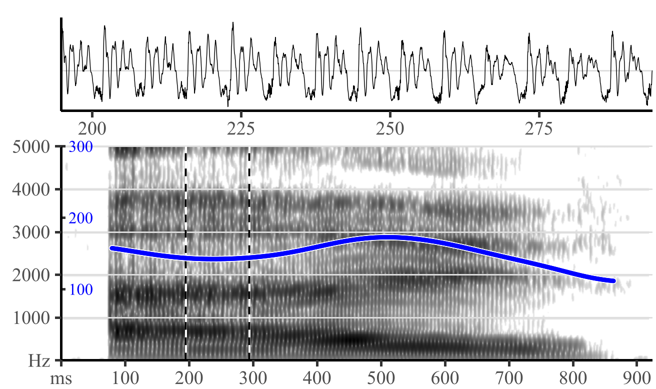

FIG. 7. Waveform excerpt, spectrogram, and pitch track for [øy^{241}] produced with harsh voice by FZ01F.

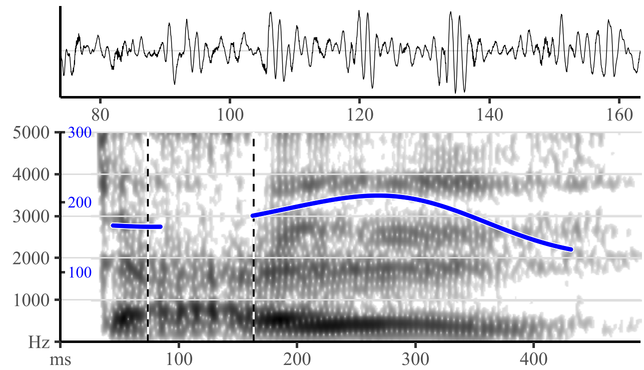

FIG. 8. Waveform excerpt, spectrogram, and pitch track for [tøy^{241}] with harsh voice and aryepiglottic trilling by FZ05F.

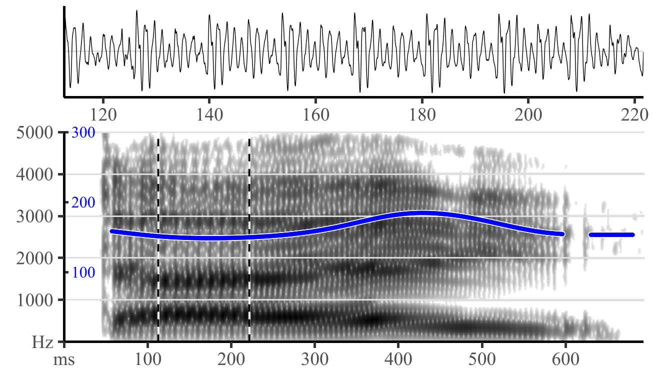

FIG. 9. Waveform excerpt, spectrogram, and pitch track for [tøy^{241}] produced with ventricular voice by FZ07F.

subharmonics spaced at multiples of 74 Hz are also present. This kind of voice quality is expected to have lower H1–H2 and lower HNR (Keating *et al.*, 2015), as the presence of subharmonics decreases the relative prominence of the primary harmonic series.

Table VI provides the relative incidence for each voicing type in the dataset. For most speakers, harsh voice (without trilling) accounts for the largest proportion and comprises nearly 60% of the 2915 tokens. The second-most common voicing types are creaky (17.39%) and modal voice (16.43%). While harsh voice is the most common voice quality in general, some interspeaker differences are observed. Two older male speakers, 03M and 04M, use harsh voice exclusively. The oldest, 03M, is the only speaker for whom aryepiglottic trilling is observed for a majority of tokens. The youngest female speakers, 07F (born 1987) and 09F (born 1974), use creaky voice most often, although 07F is also the most frequent user of ventricular voice. Other phonation types, namely, whispery voice, harsh voice with trilling, and ventricular voice, are relatively infrequent for most speakers.

Figure 10 provides the distribution of phonation types by Set B tone category. In line with the overall distribution, harsh voice without trilling is the most common phonation type for all three. Multinomial logistic regression calculated in R with VGAM (Yee, 2015) shows that, relative to /21/, /241/ and /24/ are significantly less likely to be produced with modal voice (/241/: Est. = –1.023, $p < 0.001$; /24/: Est. = –1.212, $p < 0.001$) and significantly more likely to be produced with harsh voice (/241/: Est. = 0.281, $p < 0.05$; /24/: Est. = 0.506, $p < 0.001$). /24/ is also produced with trilled harsh voice significantly more often than /21/ (Est. = 0.971, $p < 0.001$).

These differences may be explained in part by F0, given that epilaryngeal constriction is articulatorily more compatible with low pitch than with high pitch. As shown in Fig. 11, which includes F0 measurements obtained at the time of F1 maximum, /21/ tends to have a higher F0 compared with /241/ and /24/, while /24/ has a lower F0 than /241/. A linear

TABLE VI. Phonation types associated with the Set B tones.

Speaker	Modal	Creaky	Whispery	Harsh	Harsh with trilling	Ventricular
01F ($N = 170$)	14.12% (24)	34.12% (58)	9.41% (16)	40.59% (69)[a]	—	1.76% (3)
01M ($N = 173$)	14.45% (25)	19.08% (33)	—	66.47% (115)[a]	—	—
02F ($N = 154$)	16.88% (26)	5.19% (8)	9.74% (15)	64.29% (99)[a]	0.65% (1)	3.25% (5)
02M ($N = 151$)	15.89% (24)	10.60% (16)	—	73.51% (111)[a]	—	—
03F ($N = 177$)	25.42% (45)	15.82% (28)	1.69% (3)	52.54% (93)[a]	3.95% (7)	0.56% (1)
03M ($N = 144$)	—	—	—	42.36% (61)	57.64% (83)[a]	—
04F ($N = 164$)	32.32% (53)	0.61% (1)	—	64.63% (106)[a]	1.22% (2)	1.22% (2)
04M ($N = 174$)	—	—	—	100% (174)[a]	—	—
05F ($N = 165$)	28.48% (47)	3.03% (5)	—	54.55% (90)[a]	13.94% (23)	—
05M ($N = 135$)	11.85% (16)	23.7% (32)	1.48% (2)	62.22% (84)[a]	—	0.74% (1)
06F ($N = 166$)	36.14% (60)	3.61% (6)	1.20% (2)	57.83% (96)[a]	—	1.20% (2)
06M ($N = 174$)	6.32% (11)	0.57% (1)	—	91.95% (160)[a]	1.15% (2)	—
07F ($N = 166$)	33.13% (55)	33.73% (56)[a]	—	22.29% (37)	—	10.84% (18)
07M ($N = 166$)	—	36.14% (60)	0.6% (1)	58.43% (97)[a]	—	4.82% (8)
08F ($N = 152$)	1.32% (2)	22.37% (34)	—	73.03% (111)[a]	2.63% (4)	0.66% (1)
09F ($N = 161$)	6.21% (10)	75.16% (121)[a]	—	16.77% (27)	—	1.86% (3)
10F ($N = 147$)	1.36% (2)	31.29% (46)	—	64.63% (95)[a]	1.36% (2)	1.36% (2)
11F ($N = 176$)	44.89% (79)	1.14% (2)	0.57% (1)	52.84% (93)[a]	—	0.57% (1)
Total ($N = 2915$)	16.43% (479)	17.39% (507)	1.37% (40)	58.94% (1718)[a]	4.25% (124)	1.61% (47)

[a]The most frequent phonation type for each speaker.

https://doi.org/10.1121/10.0036256

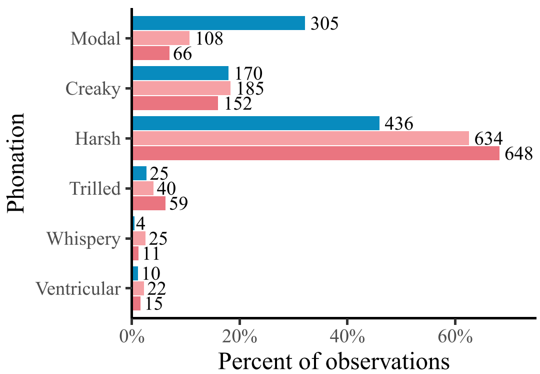

FIG. 10. Distribution of Set B phonation types by tone with token counts indicated.

mixed effects regression model was constructed to test these differences, using the following formula: `F0 ~ tone + (1 + tone|speaker) + (1|word)`. Pairwise *post hoc* comparison tests confirm that tonal F0 differences are highly significant (/21/ vs /24/: $p < 0.0001$; /21/ vs /241/: $p < 0.0001$; /24/ vs /241/: $p < 0.001$).

Finally, note that in all examples given in Figs. 3–9, the period of non-modal voicing occurs during the initial interval of the vowel's duration, while the remainder of the rhyme is produced with modal voicing. The period of non-modal voicing also typically corresponds to the lowered nucleus of diphthongized vowels (e.g., the [e] phase of the [ei] allophone for /i/), with voicing becoming modal along the pitch rises for the tones /241/ and /24/. While /21/ exhibits a pitch fall, rather than rise, constricted phonation likewise occurs during the first half of the rhyme, as can be seen in Fig. 5. It does not appear to be the case, therefore, that constricted phonation is an automatic consequence of low F0 (compare Kuang, 2017); however, more extreme types of constricted phonation are more likely to occur with tones exhibiting a lower pitch, consistent with predictions of the LAM (Esling *et al.*, 2019).

B. Acoustic correlates of epilaryngeally constricted phonation

Figure 12 provides H1*–H2* and CPP measures for the Set B tones according to phonation type. Both creaky voice and ventricular voice have slightly lower H1*–H2*, while whispery voice tends to have higher H1*–H2* than modal voice. These trends are in accordance with the predictions given in the previous section. Because whispery voice is similar to breathy voice in exhibiting a relatively wide glottal aperture, it is expected to have a higher H1*–H2*. Harsh, ventricular, and whispery phonation tends to have lower CPP values in comparison with creaky and modal phonation, while creaky voice has a slightly higher CPP than modal voice. Because CPP (and HNR) are inverse measures, lower values correspond to noisier phonation types, which involve greater epilaryngeal constriction. HNR and CPP reflect the amplitude difference between voicing harmonics and noise; when irregular vibration and the amplitude of turbulent noise increase as the result of greater constriction, the relative strength of the harmonics is diminished.

A linear mixed effects model was constructed to test the differences in H1*–H2* and CPP observed in Fig. 12, as well as any difference in H2*–H4*, using the following formula: `H1*–H2*/H2*–H4*/CPP ~ phonation type + (1|speaker) + (1|word)`; a random slope by speaker was not included due to singular fit. Pairwise *post hoc* comparison tests show that the slight H1*–H2* difference between modal voice and creaky voice is nearly significant (modal vs creaky: $p = 0.08$), while modal voice does not significantly differ in H1*–H2* from either harsh or ventricular voice (modal vs harsh: $p = 0.66$; modal vs ventricular: $p = 1.00$). Whispery voice has significantly higher H1*–H2* than modal ($p < 0.001$), creaky ($p < 0.001$), harsh ($p < 0.01$), and ventricular ($p < 0.01$) voice, while harsh voice has significantly higher H1*–H2* than creaky voice ($p < 0.001$). H2*–H4*, by contrast, does not reliably distinguish any Set B phonation type (for modal vs whispery, modal vs ventricular, creaky vs harsh, $p = 1.00$; for whispery vs ventricular, harsh vs ventricular, $p = 0.99$; for modal vs creaky, $p = 0.30$; for modal vs harsh, $p = 0.21$; for creaky vs whispery, $p = 78$; for creaky vs ventricular, $p = 0.98$; for whispery vs harsh, $p = 0.82$).

Modal voice has a significantly lower CPP than creaky voice ($p < 0.05$), and a higher CPP than all other phonation types (modal vs whispery, $p < 0.001$; modal vs harsh,

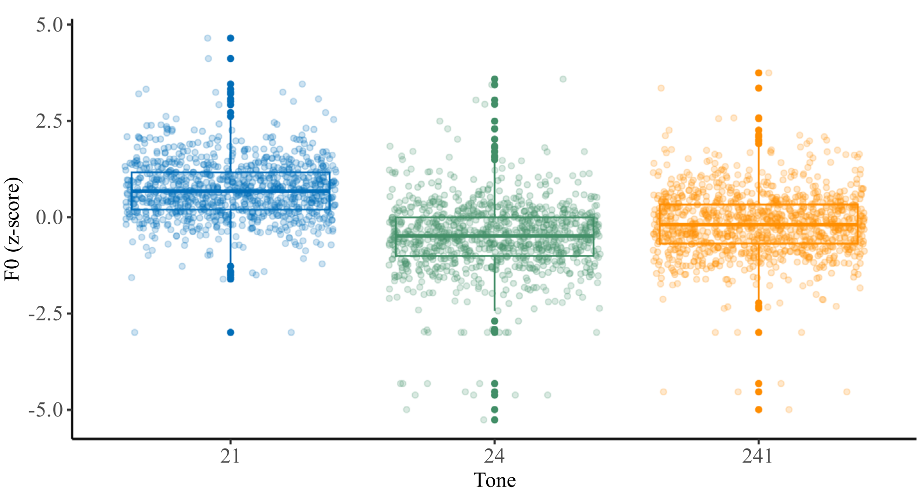

FIG. 11. F0 by tone for the Set B tones /21, 241, 24/.

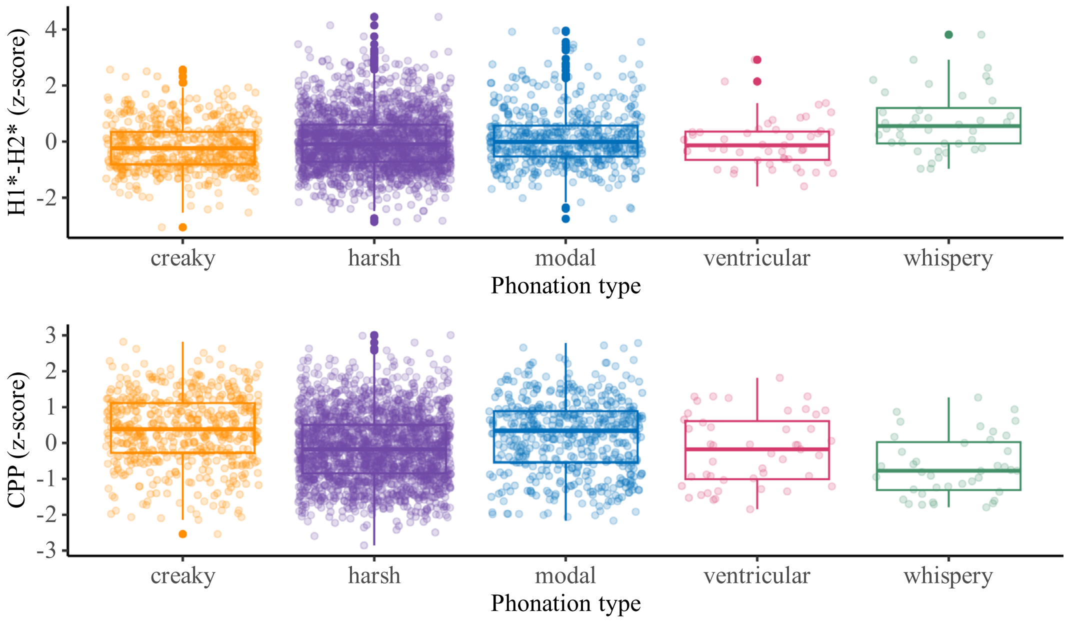

FIG. 12. H1–H2 and CPP measures for Set B tones produced with differing types of epilaryngeally constricted phonation.

$p < 0.001$; modal vs ventricular, $p < 0.05$). As expected, creaky voice has a higher CPP than whispery, harsh, and ventricular voice qualities (creaky vs whispery, $p < 0.001$; creaky vs harsh, $p < 0.001$; creaky vs ventricular, $p < 0.001$). Compared with harsh voice and ventricular voice, whispery voice has significantly lower CPP (whispery vs harsh, $p < 0.05$; whispery vs ventricular, $p < 0.05$). There is no significant difference between harsh and ventricular voice qualities ($p = 1.00$).

The spectral tilt and noise measures that best differentiate the Set B tones /21, 241, 24/ from the Set A tones /44, 51, 32/ were determined through direct comparison of Set A and Set B words with phonetically similar vowel heights. The high level Set A tone /5/ was excluded from analysis due to its very short duration, around 25% of the duration for tones in CV syllables. This difference is apparent in Fig. 1, which is consistent with quantitative duration and F0 measures reported by Peng (2011) and Donohue (2013). In total, three Set A and three Set B tones are compared. The target words included in this analysis are listed in Table X (see the Appendix).

For the Set A tones, words containing the vowels /e o a/ were examined rather than words containing /i u a/. Because /a/ does not alternate as the result of tone-vowel interaction (Peng, 2011), words containing /a/ can be straightforwardly compared between Set A and Set B. Although the Set B allophones [ei ou] belong to the same phonemic category as [i u], comparison to the vowels /e o/ allows for the relationship of vowel quality with voice quality to be distinguished from differences in voice quality inherent to the two categories of tone.

Measures of spectral tilt for each class of tone are given in Fig. 13, which suggests that spectral tilt does not clearly distinguish the two categories. Likelihood ratio tests comparing linear mixed effects models fit with and without the fixed effect of tone confirm that spectral tilt for the Set A and Set B tones does not significantly differ for most measures (H1*–H2*: $\chi^2(1) = 0.5$, $p = 0.48$; H1*–A1*: $\chi^2(1) = 1.77$, $p = 0.18$; H1*–A2*: $\chi^2(1) = 0.12$, $p = 0.73$; H1*–A3*: $\chi^2(1) = 0.07$, $p = 0.80$; H4*–H2K*: $\chi^2(1) = 0.63$, $p = 0.43$; H2K*–H5K*: $\chi^2(1) = 1.37$, $p = 0.24$). Because the Set B tones exhibit a range of within-category variability and may have either higher or lower spectral tilt than the modal Set A tones, spectral tilt does not, in general, differentiate the class of Set B tones. Set B tones have a significantly lower H2*–H4* ($\chi^2(1) = 4.83$, $p < 0.05$), which can be an acoustic correlate of constricted creaky voice (Garellek, 2019), although the difference is rather weak.

In contrast, Fig. 14 reveals that all noise measures (CPP and HNR) are lower for the Set B tones than for the Set A tones, confirming that the Set B tones exhibit greater epilaryngeal constriction than the Set A tones. Inclusion of the fixed effect for tone significantly improves model fit for all five noise measures (CPP: $\chi^2(1) = 31.24$, $p < 0.001$; HNR05: $\chi^2(1) = 28.23$, $p < 0.001$; HNR15: $\chi^2(1) = 25.10$, $p < 0.001$; HNR25: $\chi^2(1) = 21.62$, $p < 0.001$; HNR35: $\chi^2(1) = 20.53$, $p < 0.001$). Thus, in addition to differences in F0 contour, the two classes of tone are acoustically distinguishable by their relative noisiness, but not necessarily by their spectral tilt.

This conclusion is further assessed in Fig. 15, which compares the H1*–H2* and CPP measures for Set A and Set B tokens auditorily coded as exhibiting modal phonation. Here, measurements obtained at the time of F1 maximum for the Set A vowels /e, o/ are compared with tokens of the Set B vowels [ei, ou]. Speakers who produced no Set B items with modal voicing (i.e., 03M, 04M, and 07M) are necessarily excluded. In total, 262 tokens of Set A tones are compared with 338 tokens of Set B tones. H1*–H2* measures for the two sets of tone do not significantly differ

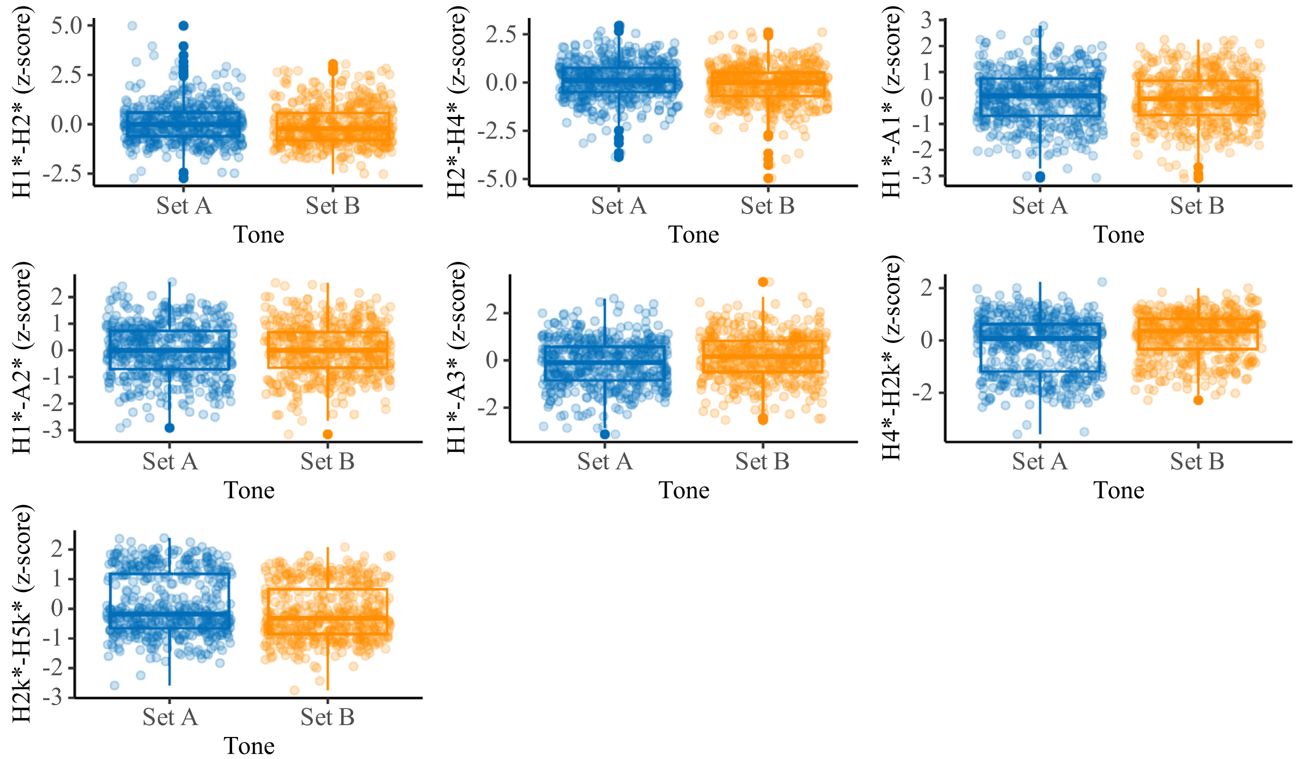

FIG. 13. Spectral tilt measures for Set A (/44, 51, 32/) and Set B (/21, 241, 24/) tones, z-score normalized.

($\chi^2(1) = 0.14$, $p = 0.71$). However, this comparison reveals that Set B tones with modal voicing exhibit significantly lower CPP measures ($\chi^2(1) = 18.92$, $p < 0.001$) when compared with the phonologically modal Set A tones. Consistent with the conclusion that the Set B tones are unified by the presence of epilaryngeal constriction, greater constriction is found to be present even for Set B tokens auditorily coded as modal.

C. Relationship between voice quality and vowel quality

Increased laryngeal constriction responsible for noisy non-modal phonation types, such as harsh voice, is expected to promote a lower and more retracted tongue position, given that lingual retraction (as compared to fronting) is articulatorily predisposed by constriction. We therefore

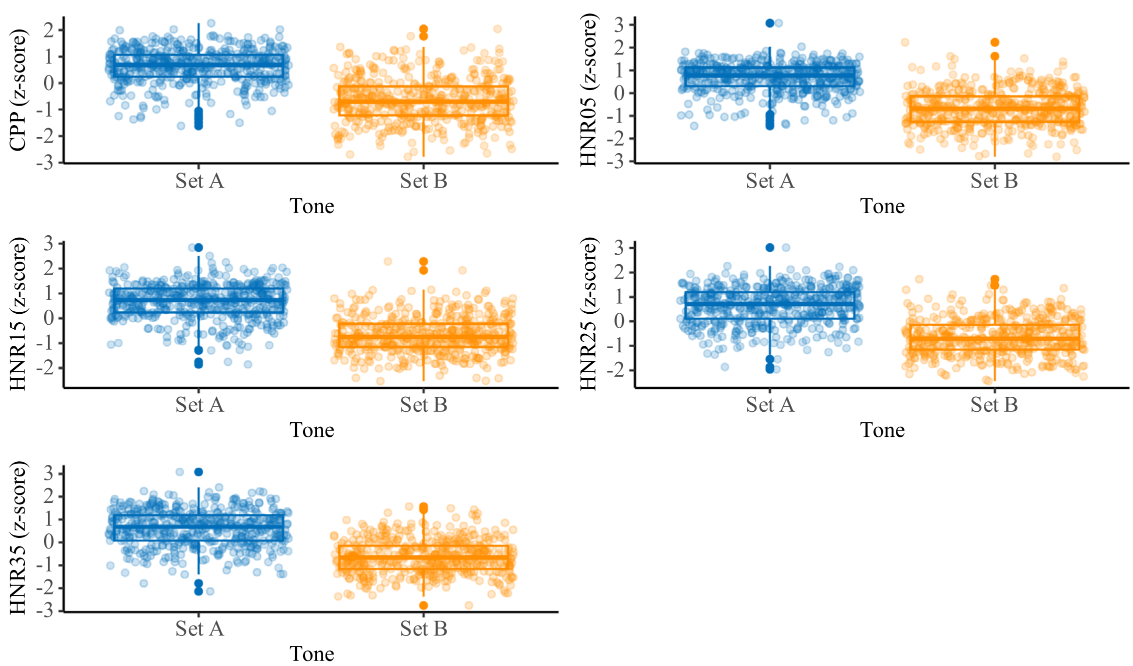

FIG. 14. Spectral noise measures for Set A (/44, 51, 32/) and Set B (/21, 241, 24/) tones, z-score normalized.

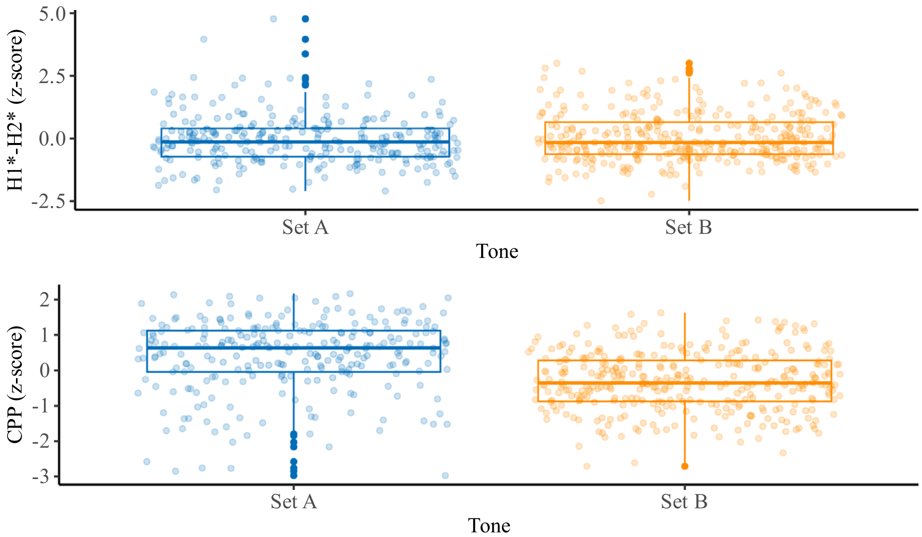

FIG. 15. H1–H2 and CPP measures for the Set A and Set B tones produced with modal voicing.

hypothesize that the formation of epilaryngeal constriction for non-modal phonation types associated with the Set B tones is responsible for variation in the degree of vowel lowering and/or retraction associated with the Set B tones. If so, we expect to observe higher F1 when the noise measure is lower (noisier). F2, on the other hand, may show vowel-specific correlations with the noise measures. Constricted vowels in Somali show an increase, decrease, or no change in F2 when compared with their non-constricted counterparts, which differs from the consistent increase to F1 for all vowels (Edmondson and Esling, 2006).

All tokens of the words given in Table XI, in addition to the Set B words given in Table X, were examined. As such, this analysis captures the entire range of phonation types observed in Sec. III A, from modal to extremely harsh. Both noise and formant measurements were extracted at the time of F1 maximum, corresponding to the lowest position achieved by the tongue throughout the rhyme. Figure 16 shows that most noise measures are negatively correlated with F1 for the vowels [ei øy ou], such that as the voice quality becomes noisier (lower CPP and HNR measures), the vowels also become lower (higher F1). While the correlation is weak for HNR05, which corresponds to the lowest frequency band of 0–500 Hz, the association is stronger for HNR measures that include progressively higher frequencies. Table VII shows that inclusion of the fixed effects for HNR15-35 improves model fit for all vowels, indicating that these measures are negatively correlated with F1. Inclusion of HNR05, on the other hand, does not significantly improve model fit for any of the vowels.

Regarding F2, [ei] generally shows a weak positive correlation with the noise measures (Fig. 17), and inclusion of HNR15 significantly improves model fit (Table VIII), indicative of a lower F2 in harsher voice due to a more retracted tongue position. On the other hand, the F2 of [ou] is negatively correlated with the noise measures (Fig. 17) and inclusion of HNR15-35 significantly improves model fit (Table VIII), suggesting that F2 increases with harsher phonation. The F2 for [øy] shows no correlation with any noise measures.

IV. DISCUSSION

This study has examined the phonetic features of non-modal phonation associated with the Fuzhounese tones /21, 241, 24/, as well as its relationship with vowel quality alternation. The use of non-modal phonation with the Set B tones has been documented in previous work (Chan, 1985; Donohue, 2011), motivating hypotheses regarding its involvement in tone-vowel interaction (Donohue, 2017; Hisao, 2010; Lam, 2014). A satisfactory link between these three features—tone, phonation, and vowel quality—has remained elusive, owing to a paucity of quantitative acoustic data and to ambiguities in the terminology used to characterize the relevant phonation types. Based on auditory observation, prior studies have described the voice quality of the Set B tones as "creaky," "glottalized," or "glottalized breathy" (Chan, 1985; Donohue, 2011, 2017; Hisao, 2010; Lam, 2014). These terms implicitly correspond to traditional models of phonation, in which phonatory quality is determined by the state of the larynx alone, including its height (Hombert *et al.*, 1979; Honda *et al.*, 1999), width of the glottal aperture (Ladefoged, 1971), and the stiffness and thickness of the vocal folds (Zhang, 2016). It is cross-linguistically common, however, for creaky voice to be produced with epilaryngeal constriction (Edmondson and Esling, 2006; Esling, 2013; Garellek, 2019; Hollien, 1972)—this supraglottal constriction can contribute to a harsh voice quality, especially when the ventricular folds or aryepiglottic folds vibrate simultaneously with the vocal folds (Esling, 2013).

Here, we find that Fuzhounese speakers produce the Set B tones with a range of phonation types, including creaky voice, ventricular voice, whispery voice, and harsh voice, the latter of which may or may not be accompanied by aryepiglottic trilling. The Set B tones are also frequently

https://doi.org/10.1121/10.0036256

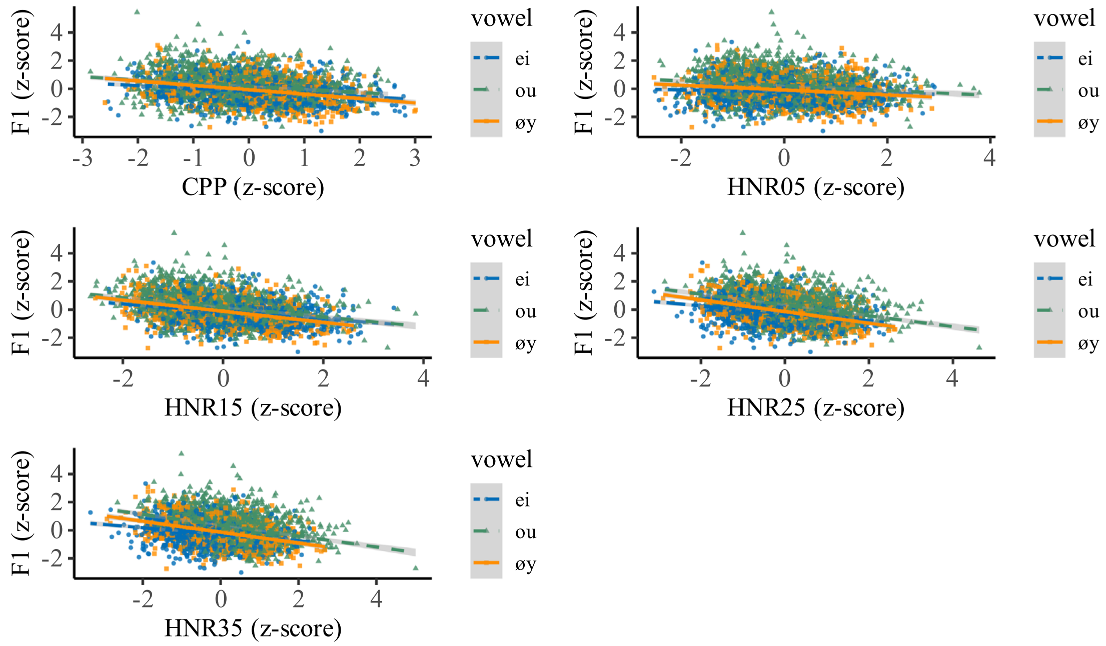

FIG. 16. Correlation of F1 with noise measures. Both formant and noise measurements are z-score normalized. Regression lines represent simple linear smooths (i.e., not including random effects).

produced with modal voice, although it was found that modal voice observed with Set B tones is acoustically distinct from, i.e., noisier than, modal voice observed for the Set A tones. Phonation type was observed to vary not only between speakers, but also on an intraspeaker basis. That is, all speakers use two or more types of constrictive non-modal phonation even across multiple repetitions of the same word. Variability observed within the Set B tones can be readily understood through the LAM (Esling *et al.*, 2019), which fundamentally differs from traditional models of phonation (Ladefoged, 1971; Zhang, 2016) and of the vocal tract more generally. Because the phonation types observed occupy nearly the entire spectrum of glottal aperture, from spread to tightly adducted, models defined by a one-dimensional continuum of vocal fold approximation do not readily account for their appearance as phonetic variants of the same phonological contrast, at least from an articulatory perspective. Such models obscure the relationship of creaky voice to harsh, whispery, and ventricular voicing, given that the latter three cannot, by definition, be achieved by manipulation of the glottis alone. Because the differences between these phonation types is not clear cut, either in articulation or acoustics (motivating psychoacoustic models) (Garellek *et al.*, 2016; Kreiman *et al.*, 2014, 2021), classifications of constricted phonation vary depending on the theoretical approach adopted. For instance, diplophonic phonation that results from entrained vibration of the ventricular folds may be regarded either as a kind of harsh voice (Esling, 2013; Laver, 1980) or as a kind of creaky voice (such as by Keating *et al.*, 2015, who refer to it as "multiply pulsed voice").

TABLE VII. ANOVA comparison between two models of F1 with and without the fixed effect of each noise measure.

	Vowel					
	[ei]		[øy]		[ou]	
Fixed effect	$\chi^2(1)$	p	$\chi^2(1)$	p	$\chi^2(1)$	p
CPP	1.77	0.18	5.79	<0.05	1.23	0.27
HNR05	1.62	0.20	0.12	0.73	0.97	0.32
HNR15	6.80	<0.01	11.83	<0.001	16.12	<0.001
HNR25	8.02	<0.01	13.34	<0.001	24.58	<0.001
HNR35	7.49	<0.01	14.57	<0.001	24.51	<0.001

The LAM postulates that the multidimensionality of phonation is better characterized as unconstricted vs constricted, rather than open vs closed. Epilaryngeal constriction unifies all phonation types observed for the Fuzhounese Set B tones, including the relatively more noisy "modal" phonation, thus reflecting a continuum of glottal and epilaryngeal tightening as well as variation in the amount of airflow (Esling, 2013). Because epilaryngeal compression involves the same articulatory mechanism as pharyngeal consonants (Esling, 1996; Esling *et al.*, 2019), the lowered allophones may be transcribed as [eˤi, øˤy, oˤu]. The use of the pharyngeal diacritic ˤ is similarly used to transcribe harsh voice by Esling (1996) and Moisik (2013), as well as by Rose (2015) for Wu Chinese and Traill (1985, 1986) for !Xóõ. With extreme epilaryngeal compression, the aryepiglottic folds are raised toward the epiglottis, allowing for aryepiglottic trilling to occur, which is likewise a feature of pharyngealization in Iraqi Arabic (Hassan *et al.*, 2011; Heselwood, 2007). When trilling is present, the Fuzhounese Set B vowels may thus be transcribed as [eʢi, øʢy, oʢu], reflecting involvement of the aryepiglottic mechanism (note the use of ʢ vs ˤ).

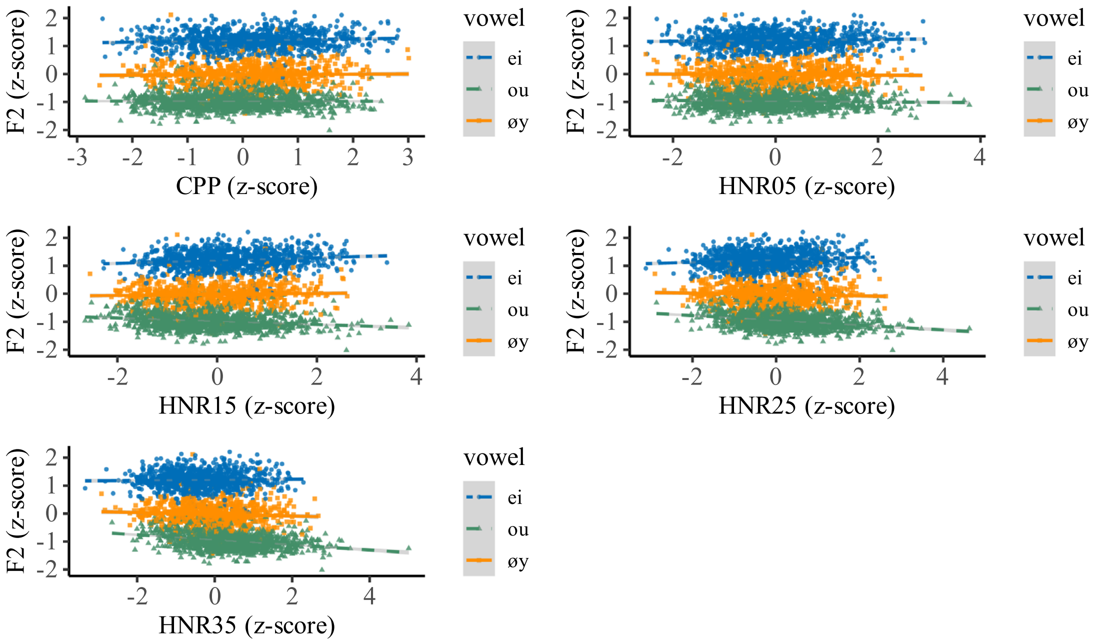

FIG. 17. Correlation of F2 with noise measures. Both formant and noise measurements are z-score normalized. Regression lines represent simple linear smooths (i.e., not including random effects).

The features observed most consistently in this study are a constricted, noisy auditory quality, so the term "harsh voice" accurately encompasses the range of Set B phonation types. The importance of noise to the realization of the Set B tones is corroborated by the results in Sec. III B. It was found that noise measures were lower for the Set B tones than for the Set A tones, demonstrating that, in addition to differences in F0, the Set B tones are most reliably defined by their noisiness. By contrast, most measures of spectral tilt did not significantly differ between the two registers, with the exception of H2*–H4*. A large body of research (introduced in Sec. II C) has found spectral tilt, in particular H1–H2, to be a reliable measure for distinguishing types of non-modal phonation in numerous languages (reviewed by Esposito and Khan, 2020). Higher H1–H2 values are understood to correspond to breathier voicing: A wider glottal aperture allows for greater airflow, reinforcing the first harmonic (Fischer-Jørgensen, 1967; Sundberg, 1979). Conversely, lower H1–H2 values are expected for creaky voice. In this study, higher H1*–H2* distinguishes whispery voice from all other phonation types, consistent with its wide glottal aperture similar to breathy voice (Esling, 2013). H1*–H2* also distinguishes harsh from creaky voice, consistent with the higher airflow required for harsh voice (Esling *et al.*, 2019), in some cases sufficient to induce aryepiglottic trilling. This measure therefore reliably distinguishes between different subcategories of phonation for the Set B tones, which exhibit disparate modes of vocal fold (and ventricular fold) vibration.

TABLE VIII. ANOVA comparison between two models of F2 with and without the fixed effect of each noise measure.

	Vowel					
	[ei]		[øy]		[ou]	
Fixed effect	$\chi^2(1)$	p	$\chi^2(1)$	p	$\chi^2(1)$	p
CPP	2.65	0.10	0.02	0.88	0.00	0.98[a]
HNR05	1.14	0.29[a]	0.49	0.48	0.55	0.46[a]
HNR15	4.42	<0.05	0.75	0.39	13.33	<0.001
HNR25	2.51	0.11	1.19	0.27	21.00	<0.001
HNR35	0.26	0.61[a]	2.17	0.14	19.69	<0.001

[a]Models fit without a random speaker slope.

The finding that H1*–H2* captures differences between some phonation types, but not between Set B and Set A tones in general, is also consistent with prior work. It has been noted that H1–H2 is less reliable for distinguishing creaky voice from modal voice (Chai and Garellek, 2022) and that it sometimes succeeds only for certain speakers, tones, or contexts. H1–H2 is a more successful measure of phonation type for high tones than mid or low tones in Mpi (Blankenship, 2002), for female than for male speakers of Santa Ana del Valle Zapotec (Esposito, 2010), at tone onset rather than offset in Jalapa Mazatec (Garellek and Keating, 2011), and in specific prosodic positions in Burmese (Gruber, 2011). Garellek and Esposito (2023) report that CPP is a more consistent correlate of non-modal phonation than H1*–H2* in White Hmong, as also found here. They attribute this result, in part, to within-corpus variability and the possibility of errors in harmonic and formant estimation. They further suggest that phonation type itself may not be the primary articulatory target for the contrast, which may instead be low F0 (creakier) and aspiration (breathier). In Fuzhounese, which also exhibits substantial variability, phonation type differences relating to vocal fold thickness or

https://doi.org/10.1121/10.0036256

aperture are apparently secondary to the primary target of epilaryngeal constriction.

H2–H4 is a somewhat newer measure (introduced by Kreiman *et al.*, 2007) and is less well understood, in particular with respect to its articulatory correlates. H2–H4 is an acoustic correlate of breathy voicing in Gujarati (Khan, 2012), although it does not correspond to breathy or creaky phonation in White Hmong (Esposito, 2012), lax phonation in Yi (Kuang, 2013), or breathy phonation in Shanghainese (Tian and Kuang, 2021). Previous work shows that H2–H4 is, at best, weakly correlated with H1–H2 (Esposito, 2012), and that both measures can independently contribute to the percept of breathiness (Garellek *et al.*, 2013). In contrast to H1–H2 (Holmberg *et al.*, 1995), H2–H4 is not correlated with proportion of glottal opening (Esposito, 2012). Garellek (2019) hypothesizes, based on modeling results from Zhang (2016), that spectral tilt measures in all frequency bands (including H2–H4) may be negatively correlated with increased vocal fold thickness, as would be expected to occur during constricted phonation. The present study finds that the Set B tones exhibit a significantly lower H2*–H4* than the Set A tones, although the difference is weak. In addition, H2*–H4* was found not to differ among any of the Set B phonation types. Although these results hold even with F0 included as a fixed effect in the linear models, it is worth noting that Kuang (2013) reports H2–H4 to differ between Yi tones distinguished by pitch (mid vs low), but not between tense and lax phonation types. Because phonation type and tone cannot be freely combined in Fuzhounese, the relative influences of F0 and laryngeal constriction are not easily disentangled. Varying results for spectral slopes at different frequency ranges would be compatible with the assumptions of psychoacoustic models, which posit that these cues may independently contribute to voice perception (Garellek *et al.*, 2016; Kreiman *et al.*, 2014). At the same time, few previous studies have examined acoustic voice quality measures in terms of the articulatory framework of the LAM and vice versa. Further research is needed to understand the relationship of such measures to both glottal and supraglottal articulation, allowing for models well grounded in both perceptibility and articulation.

Most phonological theories assume strict separation of the tongue and the larynx, which has impeded a clear explanation for the concomitant alternation of tone, phonation, and vowel quality in Fuzhounese. Rejecting this stance, the LAM straightforwardly explains phonological patterns involving segmental and suprasegmental interaction, which are argued to result from the articulatory link of the tongue to the larynx via the epilarynx. Analysis of the relationship between noise measures and formant frequencies revealed that more strongly constricted phonation (i.e., lower HNR 15–35) is significantly associated with lower vowel quality (i.e., higher F1) for all vowels, and associated with a retracted vowel quality (i.e., lower F2) for the front vowel [ei]. It was further observed that the lowest HNR band (HNR05) showed no correlation with F1 or F2 for any vowel. This can be explained by the sensitivity of HNR05 to irregular F0 (Keating *et al.*, 2015), given that this HNR band includes only frequencies below 500 Hz and excludes the expected frequency bands for frication noise generated in the pharynx (around 1500 and 2800 Hz as shown for Arabic by Johnson, 2011). As such, low HNR05 values are not necessarily attributable to tongue retraction or lowering.

The relationship of F2 with noise measures was found to vary according to the vowel's backness. While the front vowel [ei] had a positive correlation, consistent with lingual retraction, the back vowel [ou] showed a negative correlation (higher F2 with lower HNR). An increase in F2 for lowered back vowels has previously been observed in a comparison of the /o/ allophones [o] and [ɔ] in Fuzhounese (Peng, 2011) and in comparison of the non-constricted vowels [o] and [oː] with their constricted counterparts [ɔ] and [ɔː] in Somali (Edmondson and Esling, 2006). This can be explained by asymmetry of the acoustic vowel space, such that lower back vowels tend to have higher F2 than higher back vowels. The vowel nucleus for [øy] has a more centralized quality, between [e] and [o] (in agreement with acoustic data from Peng, 2011), and shows correlation of F1, but not F2, with the noise measures. Retraction of the tongue for [ø] may be avoided to prevent merger of [øy] and [ɔy]. Results for [ou] and [øy] also indicate that a lowered tongue position alone (higher F1) is sufficient for producing harsher voice through the formation of a more constricted epilaryngeal tube, which can be corroborated by future articulatory studies utilizing laryngoscopy, real-time MRI, and lingual and/or laryngeal ultrasonography. Such work will not only confirm results inferred here by acoustic measurement, but can also reveal the relative contributions of various vocal tract actions to laryngeal constriction. These include laryngeal raising, lingual retraction, and lowering of the jaw, the latter having been suggested by Hisao (2010) as a contributor to vowel quality alternation.

Vowel-specific correlations between noise measures and formant frequencies are consistent with patterns observed in Chen's (1998) cross-dialectal comparison. We find that [ei] tends to be lower and more centralized, in line with alternation between [ei] and [ai] in Fuzhounese (only in (C)Vʔ and (C)Vŋ syllables) as well as the alternation between [i] and [æi] in Minqing (Chen, 1998). [øy] shows a tendency to be lower but not retracted, in agreement with the observed alternation between [y] and [œ/øy] in Xiao'ao (Chen, 1998). Finally, the lowering and centralization of [ou] are consistent with descriptions of alternation between [ou] and [ɒu/au] in Fuzhounese and between [u] and [ʌu] in Minqing (Chen, 1998). For Fuzhounese in particular, a range of transcriptions for the lowered allophones may be found, as shown in Table IX, consistent with the variability demonstrated here. This study suggests that such variation is conditioned by degree of epilaryngeal constriction: when the Set B tones are produced with stronger constriction, vowels tend to be lower (open-mid [ɛi œy ɔu] or even lower) and phonation is noisier.

Because the gradient phonetic effects of epilaryngeal constriction are in line with the direction of historical vowel

TABLE IX. Transcriptions of Fuzhounese vowel allophones in previous work.

	Allophones of /i y u/	Phonemic /e ø o/
Tao (1930)	[ei øy ou]	[ɛ œ o]
Liang (1986)	[ɛi øy ou]	[ɛ œ o]
Feng (1998)	[ɛi øy ou]	[ɛ ø o]
Chen (1998)	[ei øy ou]	[ɛ œ o]
Yuan (2001)	[ei øy o̜u]	[ɛ œ ɔ]
Hisao (2010)	[ɛi œy ɔu]	[E ø o][a]

[a][E] is a non-IPA symbol for a front vowel between [e] and [ɛ].

lowering, synchronic variation observed in Fuzhounese and related varieties provides a window into the development of tone-vowel alternation in Eastern Min. Diachronic evidence suggests that vowel quality alternation was first observed for the high vowels /i y u/ (Maclay and Baldwin, 1870), while alternation for the mid vowels is reported later (Tao, 1930). Such a progression is consistent with the predictions of the LAM and the closely related phonological potentials model (Moisik *et al.*, 2021). Lingual fronting, as required to produce /i y/, often co-occurs with laryngeal lowering, with both actions being antagonistic (anti-synergistic) to epilaryngeal constriction (Esling *et al.*, 2019; Moisik *et al.*, 2021). Low vowels, which do not alternate even in present-day Fuzhounese (Peng, 2011), are synergistic with laryngeal constriction and are generally perceived to be harsher (Rees, 1958). Moisik *et al.* (2021) point out that a number of languages with constricted vowel registers therefore exhibit asymmetrical inventories in which high front vowels are absent from the constricted/pharyngealized series, while back and low vowels are more consistently represented (Northern Horpa: Chiu and Sun, 2020; Khoisan: Hess, 1998; Miller, 2007; Miller *et al.*, 2009; Moisik, 2013).

Cross-dialectal variability reported by Chen (1998) strongly motivates future studies comparing structurally similar varieties representing different stages in the phonologization of constricted register. As was shown in Table III, several Eastern Min varieties lack tone-vowel alternation (e.g., Luoyuan), while others (e.g., Minqing) exhibit more extreme vowel lowering than in Fuzhounese. A detailed phonetic analysis of the tone, vowel quality, and (crucially) phonation types present in those varieties will allow for a more complete, phonetically grounded explanation.

Hypothetically, interaction of tone and prosodic prominence could lead to the independent emergence of constricted phonation. Voice quality is known to interact with stress and pitch accent (Basbøll, 2005; Bird and Garellek, 2019; Fischer-Jørgensen, 1989; Garellek and White, 2015), and several prior accounts of Fuzhounese formalize the restriction of Set B tones to stressed or heavy domain-final syllables (Chan, 1985; Jiang-King, 1996; Wright, 1983). If so, anti-synergy of high vowels with epilaryngeal constriction could plausibly lead to a bias against their appearance in this position, motivating lowering and retraction. If epilaryngeally constricted phonation is found in varieties which lack phonologized tone-vowel alternation, and therefore retain high(er) vowels, we would then expect to observe phonetic variability similar to Fuzhounese. That is, more constricted phonation would be correlated with lower vowel height, supporting the hypothesis that articulation of laryngeal quality is responsible for vocalic lowering.

An alternative account, in which constricted phonation occurs as the by-product of vowel lowering, would not be expected. Experimental data bearing on this question are somewhat limited but suggest that the effects of vowel quality on phonation are not automatic, in line with the assertion that "tongue retraction biases but is not sufficient for active epilarynx stricture" (Esling *et al.*, 2019, p. 164). Koenig and Fuchs (2021) report electroglottographic data demonstrating that tense-lax vowel pairs in German, which are typically distinguished by spectral quality differences, do not systematically differ in glottal contact quotient. In an examination of acoustic data from eight diverse languages with contrastive phonation, Esposito *et al.* (2021) find that vowels with higher F1 and F2 tend to have lower $H1^*–H2^*$ (suggesting creakier phonation), while vowels with lower F1 and F2 tend to have higher $H1^*–H2^*$. Within individual vowel categories, however, vowel quality did not vary according to the degree of non-modal phonation. The latter result contrasts with results of the present study, as well as previous work on Green Mong (Andruski and Ratliff, 2000) and Yi (Kuang, 2011), but is in line with work on White Hmong (Esposito, 2012) and Gujarati (Khan, 2012). In Sec. III B, the Set A and Set B tones were shown to differ in their noise characteristics by way of comparisons made between vowel tokens with phonetically similar heights. That is, words with Set B tones contained the allophones [ei ou a], while words with Set A tones contained the phonemic vowels [e o a], yet only the Set B tones are realized with constricted phonation. It is not only the case that there is no clear phonetic motivation for vowel lowering to occur on its own. In that scenario, it would also be expected that independent changes to vowel quality would be evenly distributed across tonal categories, rather than being limited to a specific register, and vowel quality would not be expected to alternate according to tone sandhi rules.

In present-day Fuzhounese, where constricted register has been phonologized, it is likely that all three cues (pitch, phonation, and vowel quality) are perceptually relevant. Donohue (2011) reports that constricted phonation alone can perceptually distinguish the tones /21/ and /32/, which have highly similar pitch contours. Because the relationship of constricted phonation to tone and vowel quality alternation had not yet been proposed, the perceptual roles of vowel quality and phonatory quality, both of which are gradiently variable, is not yet known. The observed three-way interaction of pitch, phonation, and vowel quality therefore motivates the need for additional perceptual studies, especially because such cues may engage in trading relations and their relative importance can changeover time. Indeed, Kuang and Cui (2018) demonstrate that cues associated with the register contrast in Southern Yi (Tibeto-Burman) are undergoing change. Younger speakers rely more on vowel quality differences than phonatory differences in perceiving and producing the contrast, similar to a process also reported for the Austroasiatic language Chru (Brunelle *et al.*, 2020).

For this reason, the potential for change in Fuzhounese merits consideration through apparent-time analysis of a larger, socially balanced sample of speakers. Although a small number of speakers born later than 1970 are represented here, it is noted that the youngest women, i.e., 07F and 09F, are somewhat more likely to produce Set B tones with less constricted phonation, i.e., modal or creaky voice. By contrast, two of the older men, 03M and 04M, use highly constricted harsh voice exclusively. Whether this is reflective of a broader trend remains to be seen, but would be consistent with Lam and Hong's (2016) finding that younger female speakers of Fuqing Chinese exhibit more extreme vowel lowering. If such a process is taking place in Fuzhounese, or has already taken place in Minqing Chinese, more extreme lowering may reflect realignment from a contrast primarily signaled by phonatory quality to one signaled by vowel quality. Sociophonetic study with collection of spontaneous conversational speech would allow for a greater range of variability to be observed than was possible here. Importantly, a larger corpus would include multisyllabic utterances in a range of phrasal contexts, which are not often a focus of voice quality research (compare Garellek and Esposito, 2023). In this case, such data are essential for a full understanding of the phonological function of constricted phonation and its potential relationship to word- and phrase-level prominence. While a full phonological treatment is beyond the scope of the present study, we submit that epilaryngeal constriction offers the strongest potential for a grounded account of tone-vowel interaction, with several promising avenues for future research.

V. CONCLUSION

This study has shown that the Fuzhounese tones /21, 241, 24/ are usually produced with harsh voice (sometimes with aryepiglottic trilling) or creaky voice, both of which involve epilaryngeal constriction. The degree of constriction is variable, also yielding modal, whispery, and ventricular voice, so noise measures (along with F0 and vowel quality) most robustly differentiate the constricted Set B tones /21, 241, 24/ from the unconstricted Set A tones /44, 51, 32, 5/. Correlation of noise measures with F1 and F2 supports the hypothesis that tone-vowel interaction is mediated by the laryngeal quality associated with /21, 241, 24/, resulting from the articulatory affinity of lingual retraction and laryngeal raising with epilaryngeal constriction. These findings therefore suggest that lowered vowel allophones observed throughout Eastern Min occur as the byproducts of the laryngeal articulation of constricted vocal quality, which may be corroborated by future articulatory, perceptual, sociophonetic, and cross-dialectal studies. Most prior research on Fuzhounese phonology has overlooked phonation as a mediating factor in tone-vowel interaction, in large part due to a lack of clarity surrounding its phonetic characteristics, noted in fact by Chan (1985, p. 180). A lack of attention paid to phonation is far from limited to Fuzhounese (see Moisik *et al.*, 2021), illustrating the importance of theoretical mechanisms (such as the LAM) that allow for more accurate identification and formalization of lingual-laryngeal interactions. This study builds on a growing body of research demonstrating the inseparability of laryngeal and supralaryngeal articulation (Esling *et al.*, 2019; Moisik *et al.*, 2021), as well as the integral nature of voice quality research to phonetic and phonological theory (Garellek, 2022).

ACKNOWLEDGMENTS

This work was supported by a Louis Cha Postgraduate Research Fellowship awarded to C.C. We thank associate editor Jianjing Kuang and three anonymous reviewers for their valuable comments on earlier versions of the paper.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no competing interests to declare.

This study was approved by the University of Hong Kong Human Research Ethics Committee (Reference No. EA210377).

DATA AVAILABILITY

Data will be made available on request to the authors.

APPENDIX: WORD LIST

The experiment word list is given in Tables X and XI.

TABLE X. Words chosen for analysis in Sec. III B. Words with Set B tones were also analyzed in Sec. III C.

	Set A			Set B	
	IPA	Gloss		IPA	Gloss
低	[te^{44}]	low	滴	[teiʔ^{24}]	drop
題	[te^{51}]	topic	地	[tei^{241}]	land
底	[te^{32}]	bottom	蒂	[tei^{21}]	stalk
多[a]	[to^{44}]	many	涿[a]	[touʔ^{24}]	to point
玻	[po^{44}]	glass	腐	[pou^{241}]	to rot
婆	[po^{51}]	old woman	腹	[pouʔ^{24}]	belly
寶	[po^{32}]	treasure	富	[pou^{21}]	rich
疤	[pa^{44}]	scar	百	[paʔ^{24}]	hundred
爬	[pa^{51}]	to crawl	下	[a^{241}]	beneath
飽	[pa^{32}]	full	霸	[pa^{21}]	dominant
			咬[a]	[ka^{241}]	to bite

[a]Words included for some speakers as alternatives for items with variable pronunciation.

TABLE XI. Set B words entered into the analysis in Sec. III C (along with Set B words given in Table X).

	IPA	Gloss		IPA	Gloss
痹	[pei^{21}]	paralysis	鼻	[pei^{241}]	to smell
筆	[peiʔ^{24}]	pen	痣	[t͡sei^{21}]	mole
字	[t͡sei^{241}]	character	積	[t͡seiʔ^{24}]	to amass
四	[sei^{21}]	four	是	[sei^{241}]	to be
塞	[seiʔ^{24}]	to stuff	記	[kei^{21}]	to memorize
忌	[kei^{241}]	jealous	急	[keiʔ^{24}]	urgent
意	[ei^{21}]	intention	味	[ei^{241}]	taste
一	[eiʔ^{24}]	one			
膩	[løy^{21}]	to be tired of	箸	[tøy^{241}]	chopsticks
竹	[tøyʔ^{24}]	bamboo	肆	[søy^{21}]	rotten
事	[søy^{241}]	matter	宿	[søyʔ^{24}]	veteran
漬	[t͡søy^{21}]	to soak	炷	[t͡søy^{241}]	CLF (incense)
粥	[t͡søyʔ^{24}]	congee	鋸	[køy^{21}]	saw
具	[køy^{241}]	crutch	—	[køyʔ^{24}]	to submerge
—	[øy^{21}]	to seep	譽	[øy^{241}]	worn
吐	[tʰou^{21}]	to vomit	肚	[tou^{241}]	abdomen
出	[t͡sʰtʰouʔ^{24}]	to exit	露	[lou^{21}]	dew
鹵	[lou^{241}]	sauce	—	[louʔ^{24}]	to slip
數	[sou^{21}]	number	助	[t͡sou^{241}]	to help
束	[souʔ^{24}]	restraint	故	[kou^{21}]	former
顧	[kou^{21}]	to care for	褲	[kʰou^{21}]	trousers
舊	[kou^{241}]	old	—	[kouʔ^{24}]	to get up
副	[hou^{21}]	auxiliary	戶	[hou^{241}]	household
護	[hou^{241}]	to protect	福	[houʔ^{24}]	good fortune
拂	[houʔ^{24}]	to whisk	有	[ou^{241}]	to have
熨	[ouʔ^{24}]	to iron			

[1]Many sources (e.g., Chan, 1985; Donohue, 2013; Jiang-King, 1996; Yip, 1980) transcribe /5/ and /24/ with an underline (i.e., /5, 24/) to indicate their shorter duration. Both underlining and the use of a single digit (as in /5/ vs /55/) are common conventions in Chinese phonetics for transcribing checked tones.

Andruski, J. E., and Ratliff, M. (**2000**). "Phonation types in production of phonological tone: The case of Green Mong," J. Int. Phon. Assoc. **30**(1), 37–61.

Bailly, L., Henrich, N., and Pelorson, X. (**2010**). "Vocal fold and ventricular fold vibration in period-doubling phonation: Physiological description and aerodynamic modeling," J. Acoust. Soc. Am. **127**(5), 3212–3222.

Basbøll, H. (**2005**). *The Phonology of Danish* (Oxford University Press, Oxford, UK).

Bates, D., Mächler, M., Bolker, B., and Walker, S. (**2015**). "Fitting linear mixed-effects models using lme4," J. Stat. Softw. **67**(1), 1–48.

Bauer, R. S., and Benedict, P. K. (**1997**). *Trends in Linguistics Studies and Monographs Modern Cantonese Phonology* (de Gruyter, Berlin, Germany).

Bird, E., and Garellek, M. (**2019**). "Dynamics of voice quality over the course of the English utterance," in *Proceedings of the 19th International Congress of Phonetic Sciences*, edited by S. Calhoun, P. Escudero, M. Tabain, and P. Warren (Australasian Speech Science and Technology Association Inc., Canberra, Australia).

Blankenship, B. (**2002**). "The timing of nonmodal phonation in vowels," J. Phon. **30**(2), 163–191.

Boersma, P., and Weenink, D. (**2024**). "Praat: Doing phonetics by computer [computer program]."

Brunelle, M. (**2005**). "Register in Eastern Cham: Phonological, phonetic and sociolinguistic approaches," Ph.D. thesis, Cornell University, Ithaca, NY.

Brunelle, M., and Kirby, J. (**2016**). "Tone and phonation in Southeast Asian languages," Lang. Linguist. Compass **10**(4), 191–207.

Brunelle, M., Tạ, T. T., Kirby, J., and Giang, Đ. L. (**2020**). "Transphonologization of voicing in Chru: Studies in production and perception," Lab. Phonol. **11**(1), 15.

Chai, Y., and Garellek, M. (**2022**). "On H1–H2 as an acoustic measure of linguistic phonation type," J. Acoust. Soc. Am. **152**(3), 1856–1870.

Chan, M. K. M. (**1985**). "Fuzhou phonology: A non-linear analysis of tone and stress," Ph.D. thesis, University of Washington, Seattle, WA.

Chen, W.-R., Whalen, D., and Tiede, M. K. (**2021**). "A dual mechanism for intrinsic f0," J. Phon. **87**, 101063.

Chen, Y., and Gussenhoven, C. (**2015**). "Shanghai Chinese," J. Int. Phon. Assoc. **45**(3), 321–337.

Chen, Z. (**1998**). *Study of Fuzhou Chinese* (Fujian People's, Fuzhou, China).

Chiu, C., and Sun, J. T.-S. (**2020**). "On pharyngealized vowels in Northern Horpa: An acoustic and ultrasound study," J. Acoust. Soc. Am. **147**(4), 2928–2946.

Denning, K. (**1989**). "The diachronic development of phonological voice quality, with special reference to Dinka and the other Nilotic languages," Ph.D. thesis, Stanford University, Stanford, CA.

Donohue, C. (**2011**). "The significance of 'secondary cues' for tonal identification in Fuzhou," in *Proceedings of the 17th International Congress of Phonetic Sciences*, edited by W.-S. Lee and E. Zee (City University of Hong Kong, Hong Kong).

Donohue, C. (**2013**). *Fuzhou Tonal Acoustics and Tonology* (LINCOM Europa, Munich, Germany).

Donohue, C. (**2017**). "Tones and vowels in Fuzhou revisited," in *Segmental Structure and Tone*, edited by W. Kehrein, B. Köhnlein, P. Boersma, and M. Van Oostendorp (de Gruyter, Berlin, Germany).

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

Edmondson J. A., Esling, J. H., and Lama, Z. (**2017**). "Nuosu Yi," J. Int. Phon. Assoc. **47**(1), 87–97.

Esling, J., Fattori, S. E., Biologici, N., and Variazione, F. (**2017**). *The Laryngeal Articulator's Influence on Voice Quality and Vowel Quality*, edited by C. Bertini, C. Celata, G. Lenoci, C. Meluzzi, and I. Ricci (Officinaventuno, Lombardy, Italy), pp. 13–26.

Esling, J. H. (**1996**). "Pharyngeal consonants and the aryepiglottic sphincter," J. Int. Phon. Assoc. **26**(2), 65–88.

Esling, J. H. (**2005**). "There are no back vowels: The laryngeal articulator model," Can. J. Linguist. **50**(1), 13–44.

Esling, J. H. (**2013**). "Voice and phonation," in *The Bloomsbury Companion to Phonetics*, edited by M. J. Jones and R.-A. Knight (Bloomsbury, London, UK), pp. 110–125.

Esling, J. H., Moisik, S. R., Benner, A., and Crevier-Buchman, L. (**2019**). *Voice Quality: The Laryngeal Articulator Model* (Cambridge University Press, London, UK).

Esposito, C. M. (**2010**). "Variation in contrastive phonation in Santa Ana Del Valle Zapotec," J. Int. Phon. Assoc. **40**(2), 181–198.

Esposito, C. M. (**2012**). "An acoustic and electroglottographic study of White Hmong tone and phonation," J. Phon. **40**(3), 466–476.

Esposito, C. M., and Khan, S. u. D. (**2020**). "The cross-linguistic patterns of phonation types," Lang. Linguistics Compass **14**(12), e12392.

Esposito, C. M., Sleeper, M., and Schäfer, K. (**2021**). "Examining the relationship between vowel quality and voice quality," J. Int. Phon. Assoc. **51**(3), 361–392.

Fagan, J. L. (**1988**). "Javanese intervocalic stop phonemes: The light/heavy distinction," Stud. Austronesian Linguist. **76**, 173–202.

Faytak, M. (**2014**). "High vowel fricativization and chain shift," UC Berkeley Phonol. Lab. Annu. Rep. **10**, 52–100.

Feng, A. (**1998**). *Dictionary of Fuzhou Chinese* (Jiangsu Education Press, Nanjing, China).

Fischer-Jørgensen, E. (**1967**). "Phonetic analysis of breathy (murmured) vowels in Gujarati," Ind. Linguistics **28**, 71–139.

Fischer-Jørgensen, E. (**1989**). "Phonetic analysis of the stød in Standard Danish," Phonetica **46**(1), 1–59.

Gao, J., and Hallé, P. (**2017**). "Phonetic and phonological properties of tones in Shanghai Chinese," Cah. Linguist. Asie Orient. **46**(1), 1–31.

Gao, J., Hallé, P., and Draxler, C. (**2020**). "Breathy voice and low-register: A case of trading relation in Shanghai Chinese tone perception?" Lang. Speech **63**(3), 582–607.

Garellek, M. (**2019**). "The phonetics of voice," in *The Routledge Handbook of Phonetics*, edited by W. F. Katz and P. F. Assmann (Routledge, London, UK), pp. 75–106.

Garellek, M. (**2022**). "Theoretical achievements of phonetics in the 21st century: Phonetics of voice quality," J. Phon. **94**, 101155.
Garellek, M., and Esposito, C. M. (**2023**). "Phonetics of White Hmong vowel and tonal contrasts," J. Int. Phon. Assoc. **53**(1), 213–232.
Garellek, M., and Keating, P. (**2011**). "The acoustic consequences of phonation and tone interactions in Jalapa Mazatec," J. Int. Phon. Assoc. **41**(2), 185–205.
Garellek, M., Keating, P., Esposito, C. M., and Kreiman, J. (**2013**). "Voice quality and tone identification in White Hmong," J. Acoust. Soc. Am. **133**(2), 1078–1089.
Garellek, M., Samlan, R., Gerratt, B. R., and Kreiman, J. (**2016**). "Modeling the voice source in terms of spectral slopes," J. Acoust. Soc. Am. **139**(3), 1404–1410.
Garellek, M., and White, J. (**2015**). "Phonetics of Tongan stress," J. Int. Phon. Assoc. **45**(1), 13–34.
Ge, C., Xu, W., Gu, W., and Mok, P. P. K. (**2023**). "The change in breathy voice after tone split: A production study of Suzhou Wu Chinese," J. Phon. **98**, 101239.
Gordon, M., and Ladefoged, P. (**2001**). "Phonation types: A cross-linguistic overview," J. Phon. **29**(4), 383–406.
Gruber, J. (**2011**). "An articulatory, acoustic, and auditory study of Burmese tone," Ph.D. thesis, Georgetown University, Washington, DC.
Hanson, H. M., Stevens, K. N., Kuo, H.-K. J., Chen, M. Y., and Slifka, J. (**2001**). "Towards models of phonation," J. Phon. **29**(4), 451–480.
Hassan, Z. M., Esling, J. H., Moisik, S. R., and Crevier-Buchman, L. (**2011**). "Aryepiglottic trilled variants of /ʕ, ħ/ in Iraqi Arabic," in *Proceedings of the 17th International Congress of Phonetic Sciences*, edited by W.-S. Lee and E. Zee (City University of Hong Kong, Hong Kong), pp. 831–834.
Heselwood, B. (**2007**). "The 'tight approximant' variant of the Arabic 'ayn,' " J. Intl. Phon. Assoc. **37**(1), 1.
Hess, S. A. (**1998**). "Pharyngeal articulations," Ph.D. thesis, University of Los Angeles, Los Angeles, CA.
Hisao, H. (**2010**). "On the Tonemic system and Bianyun in the Fuzhou dialect," *Diachronic Change and Language Contact: Dialects in South East China*, edited by S. H.-N. Cheung and S. H. Chang (Chinese University Press of Hong Kong, Hong Kong), pp. 58–67.
Hollien, H. (**1972**). "Three major vocal registers: A proposal," in *Proceedings of the Seventh International Congress of Phonetic Sciences*, edited by A. Rigault and R. Charbonneau (Mouton, Montreal, Canada), pp. 320–331.
Holmberg, E. B., Hillman, R. E., Perkell, J. S., Guiod, P. C., and Goldman, S. L. (**1995**). "Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice," J. Speech Lang. Hear. Res. **38**(6), 1212–1223.
Hombert, J.-M., Ohala, J. J., and Ewan, W. G. (**1979**). "Phonetic explanations for the development of tones," Language **55**(1), 37.
Honda, K., Hirai, H., Masaki, S., and Shimada, Y. (**1999**). "Role of vertical larynx movement and cervical lordosis in F0 control," Lang. Speech **42**(4), 401–411.
Huang, Y. (**2022**). "Articulatory properties of period-doubled voice in Mandarin," in *Speech Prosody 2022, ISCA*.
Jiang-King, P. (**1996**). "An optimality account of tone-vowel interaction in Northern Min," Ph.D. thesis, Rutgers University, New Brunswick, NJ.
Johnson, K. (**2011**). *Acoustic and Auditory Phonetics*, 3rd ed. (Wiley-Blackwell, Malden, MA).
Keating, P. A., Garellek, M., and Kreiman, J. (**2015**). "Acoustic properties of different kinds of creaky voice," in *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*, edited by the Scottish Consortium for ICPhS (University of Glasgow, Glasgow, UK).
Khan, S. U. D. (**2012**). "The phonetics of contrastive phonation in Gujarati," J. Phon. **40**(6), 780–795.
Koenig, L. L., and Fuchs, S. (**2021**). "Assessing vowel effects on voice quality, and voice quality effects on the respiratory system," JASA Express Lett. **1**(2), 025204.
Kreiman, J., Gerratt, B. R., and Antoñanzas-Barroso, N. (**2007**). "Measures of the Glottal Source Spectrum," J. Speech. Lang. Hear. Res. **50**(3), 595–610.
Kreiman, J., Gerratt, B. R., Garellek, M., Samlan, R., and Zhang, Z. (**2014**). "Toward a unified theory of voice production and perception," Loquens **1**(1), e009.
Kreiman, J., Lee, Y., Garellek, M., Samlan, R., and Gerratt, B. R. (**2021**). "Validating a psychoacoustic model of voice quality," J. Acoust. Soc. Am. **149**(1), 457–465.
Kuang, J. (**2011**). "Production and perception of the phonation contrast in Yi," Master's thesis, University of California, Los Angeles, CA.
Kuang, J. (**2013**). "Phonation in tonal contrasts," Ph.D. thesis, University of California, Los Angeles, CA.
Kuang, J. (**2017**). "Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice," J. Acoust. Soc. Am. **142**(3), 1693–1706.
Kuang, J., and Cui, A. (**2018**). "Relative cue weighting in production and perception of an ongoing sound change in Southern Yi," J. Phon. **71**, 194–214.
Ladefoged, P. (**1971**). *Preliminaries to Linguistic Phonetics* (University of Chicago Press, Chicago, IL).
Lam, M. F. (**2014**). "A phonetic study of tones and vowels in Fuqing Chinese," Ph.D. thesis, Hong Kong University of Science and Technology, Hong Kong.
Lam, M. F., and Hong, Y. (**2016**). "Lax tense vowels Fuqing dialect," Dialect **3**, 316–322.
Laver, J. (**1980**). *The Phonetic Description of Voice Quality* (Cambridge University Press, Cambridge, UK).
Liang, Y. (**1986**). "Phonological alternations of Fuzhou Chinese in connected speech," Stud. Lang. Linguistics **11**, 85–97.
Lotto, A., Holt, L., and Kluender, K. (**1997**). "Effect of voice quality on perceived height of English vowels," Phonetica **54**(2), 76–93.
Maclay, R. S., and Baldwin, C. C. (**1870**). *An Alphabetic Dictionary of the Chinese Language in the Foochow Dialect* (Methodist Episcopal Mission Press, Fuzhou, China).
Maddieson, I. (**1976**). "Intrinsic pitch vowels tones in Foochow," Stud. Prod. Perception Tones Univ. California Work Papers Phon. **33**, 191–202.
Miller, A. L. (**2007**). "Guttural vowels and guttural co-articulation in Ju|'hoansi," J. Phon. **35**(1), 56–84.
Miller, A. L., Brugman, J., Sands, B., Namaseb, L., Exter, M., and Collins, C. (**2009**). "Differences in airstream and posterior place of articulation among N|uu clicks," J. Int. Phon. Assoc. **39**(2), 129–161.
Mohr, B. (**1971**). "Intrinsic variations in the speech signal," Phonetica **23**(2), 65–93.
Moisik, S. R. (**2013**). "The epilarynx in speech," Ph.D. thesis, University of Victoria, Victoria, British Columbia, Canada.
Moisik, S. R., Czaykowska-Higgins, E., and Esling, J. H. (**2021**). "Phonological potentials and the lower vocal tract," J. Int. Phon. Assoc. **51**(1), 1–35.
Moisik, S. R., and Esling, J. H. (**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**(2), S687–S704.
Peng, G. (**2011**). "A phonetic study of Fuzhou Chinese," Ph.D. thesis, City University of Hong Kong, Hong Kong.
R Core Team (**2024**). "R: A language and environment for statistical computing," https://www.R-project.org/.
Rees, M. (**1958**). "Some variables affecting perceived harshness," J. Speech Hear. Res. **1**(2), 155–168.
Rose, P. (**2015**). "Tonation in three Chinese Wu dialects," in *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*, edited by the Scottish Consortium for ICPhS (University of Glasgow, Glasgow, UK).
Rose, P. (**2020**). "Variation in spectral slope and interharmonic noise in Cantonese tones," in *Interspeech 2020, ISCA*.
Shi, R. (**1998**). "The tendency of strong friction in high vowels in Chinese varieties," Linguist. Study **34**(1), 100–109.
Shue, Y.-L., Keating, P., Vicenik, C., and Yu, K. (**2011**). "VoiceSauce: A program for voice analysis," in *Proceedings of the 17th International Congress of Phonetic Sciences*, edited by W.-S. Lee and E. Zee (City University of Hong Kong, Hong Kong), pp. 1846–1849.
Stevens, K. N. (**1977**). "Physics of laryngeal behavior and larynx modes," Phonetica **34**(4), 264–279.
Sundberg, J. (**1979**). "Waveform and spectrum of the glottal voice source," in *Frontiers of Speech Communication Research, Festschrift for Gunnar Fant*, edited by B. Lindblom and S. Öhman (Academic Press, London, UK), pp. 301–320.

Tạ, T. T., Brunelle, M., and Nguyễn, T. Q. (**2022**). "Voicing and register in Ngãi Giao Chrau: Production and perception studies," J. Phon. **90**, 101115.

Tao, Y.-M. (**1930**). "Study of (Fuzhou) Min's sounds," Bull. Nat. Res. Inst. Hist. Philol. **1**(4), 445–470.

Thurgood, E. (**2004**). "Phonation types in Javanese," Oceanic Linguist. **43**(2), 277–295.

Tian, J., and Kuang, J. (**2021**). "The phonetic properties of the non-modal phonation in Shanghainese," J. Int. Phon. Assoc. **51**(2), 202–228.

Traill, A. (**1985**). *Phonetic and Phonological Studies of Xóõ Bushman* (Buske, Hamburg, Germany).

Traill, A. (**1986**). "The laryngeal sphincter as a phonatory mechanism in Xóõ Bushman," in *Variation, Culture and Evolution in African Populations: Papers in Honour of Dr Hertha de Villiers*, edited by R. Singer and J. K. Lundy (Witwatersrand University Press, Johannesburg, South Africa), pp. 123–131.

Wang, W. S.-Y. (**1972**). "The many uses of F0," in *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre*, edited by A. Valdman (De Gruyter, Berlin, Germany), pp. 487–504.

Whalen, D., and Levitt, A. G. (**1995**). "The universality of intrinsic F0 of vowels," J. Phon. **23**(3), 349–366.

Wright, M. S. (**1983**). "A metrical approach to tone sandhi in Chinese dialects," Ph.D. thesis, University of Massachusetts, Amherst, MA.

Xu, W., and Mok, P. (**2021**). "The acoustic correlates and time span of the non-modal phonation in Kunshan Wu Chinese," in *2021 12th International Symposium on Chinese Spoken Language Processing, IEEE*.

Yee, T. W. (**2015**). *Vector Generalized Linear and Additive Models: With an Implementation in R* (Springer, New York).

Yip, M. J. (**1980**). "The tonal phonology of Chinese," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.

Yip, M. J. W. (**2002**). *Tone* (Cambridge University Press, London, UK).

Yu, K. M., and Lam, H. W. (**2014**). "The role of creaky voice in Cantonese tonal perception," J. Acoust. Soc. Am. **136**(3), 1320–1333.

Yuan, J. (**2001**). *A Survey of Major Chinese Varieties* (Language Press, Beijing, China).

Zhang, Z. (**2016**). "Cause-effect relationship between vocal fold physiology and voice production in a three-dimensional phonation model," J. Acoust. Soc. Am. **139**(4), 1493–1507.