Phonetics of White Hmong vowel and tonal contrasts

Marc Garellek[iD]

University of California, San Diego
mgarellek@ucsd.edu

Christina M. Esposito[iD]

Macalester College
esposito@macalester.edu

Hmong languages, particularly White Hmong, are well studied for their complex tone systems that incorporate pitch, phonation, and duration differences. Still, prior work has made use mostly of tones elicited in their citation forms in carrier phrases. In this paper, we provide a detailed description of both the vowel and tone systems of White Hmong from recordings of read speech. We confirm several features of the language, including the presence of nasal vowels (rather than derived nasalized vowels through coarticulation with a coda [ŋ]), the description of certain tone contours, and the systematic presence of breathy and creaky voice on two of the tones. We also find little evidence of additional intonational f0 targets. However, we show that some tones vary greatly by their position in utterance, and propose novel descriptions for several of them. Finally, we show that H1*–H2*, a widely used measure of voice quality and phonation in Hmong and across languages, does not adequately distinguish modal from non-modal phonation in this data set, and argue that noise measures like Cepstral Peak Prominence (CPP) are more robust to phonation differences in corpora with more variability.

1 Introduction

White Hmong (*Hmoob Dawb*, also known as *Hmong Daw*) is a West Hmongic language, belonging to the larger Hmong-Mien language family. Although relatively understudied from the point of view of its morpho-phonology, syntax, and semantics (see Fuller 1987, Riddle 1994, Creswell & Snyder 2000, Jarkey 2010), the phonetic properties of tone and phonation in the language are fairly well researched. This is due in part to the fact that 'language description frequently focuses on aspects of a language that are notable from an areal, typological, or theoretical perspective' (DiCanio et al. 2019: 333). Indeed, with its mixed phonation-tone system, in which breathy voice is more straightforwardly contrastive than creaky voice (Garellek et al. 2013), White Hmong is particularly well suited for investigating questions regarding register/tone systems in Southeast Asian languages, as well as the phonetics of breathy and creaky voice and the role of phonation within tone systems more generally.

Despite the numerous phonetic studies of White Hmong tone and phonation, we know extremely little about how these features are realized in broader contexts. This is because previous studies have focused on the phonetics of tone and phonation as they occur in citation forms. It is common for phonetic studies to rely on data elicited from finely-curated wordlists, maximizing control over factors such as speech rate, phrasal stress, and intonation. However,

even some of the earliest quantitative studies in phonetics more generally broadened their methodology to include more naturalistic speech styles. For example, Lisker & Abramson (1967) compared the voice onset time (VOT) of words produced in isolation versus within an utterance, finding that the VOT of stops produced in isolated words was more categorically distinct than the VOT of stops produced in utterances. Others have continued this approach, in particular with VOT, measuring this feature in more spontaneous speech, including read speech (e.g. Baran, Laufer & Daniloff 1977, Crystal & House 1988, Byrd 1993, Yao 2009, Sonderegger 2012, Sleeper 2020).

As for features like tone and phonation, we also find variation in phonetic realization as a result of position in utterance. For example, researchers have shown that tones readily coarticulate with other tones (e.g. Ratliff 1992; Gandour, Potisuk & Dechongkit 1994; Xu 1994, 1997; Peng 1997; Brunelle 2009; Chang & Hsieh 2012; DiCanio 2014; Carroll 2015), such that tonal contours may vary greatly as a function of elicitation style. Moreover, Esposito (2010) demonstrated that phonation in Santa Ana del Valle Zapotec was sensitive to prosodic position, finding that non-modal categories are most robust in position with low intonational f0; even when words were uttered in isolation, as in a wordlist, the phonation contrast was minimized, possibly due to a focused reading with a higher f0. And English glottalization (e.g. of vowel-initial words) is known to vary in degree as a function of phrasing and prominence (Pierrehumbert & Talkin 1992, Dilley, Shattuck-Hufnagel & Ostendorf 1996, Davidson & Erker 2014, Garellek 2014). Variation in phonation can also occur due to speech style. Given the variation in breathy vowels seen in spontaneous vs. more careful speech in Gujarati, Khan (2012) avoided triggering a neutralization of breathy voice in 'educated' speech by asking speakers to produce spontaneous sentences containing target words.

The focus of the current paper is to provide an in-depth acoustic phonetic description of the phonological contrasts of White Hmong vowels as they occur in a corpus of read speech. We include descriptions of both vowel quality contrasts, as well as tonal contrasts, which are realized acoustically via changes in pitch, voice quality, and duration. The combined role that acoustic parameters such as formants, f0, voice quality, and duration may play in differentiating various vocalic and suprasegmental contrasts in the language is largely underexplored, particularly in more naturalistic speech. In the remainder of this section, we provide information about the language and sound inventory, focusing in particular on the vowel and tonal contrasts. The goal of this study is to further phonetic description of the language, and as such is not motivated by particular theoretical questions. Consequently, we avoid statistical comparisons across categories, focusing instead on summary statistics (in particular, means and standard deviations) of each relevant category. However, when the description is pertinent to theoretical issues, these will be highlighted in the discussion.

1.1 Language background

There are approximately 1,700,400 speakers of White Hmong (*Ethnologue*: https://www.ethnologue.com/language/MWW, accessed 8 June 2021), with most speakers living in China, Vietnam, Laos, Thailand, Burma and, via migration from Laos, the United States and other Western countries. It is closely related to the Green Mong variety (*Moob Lees*, *Moob Ntsuab*, also known as *Mong Leng* or *Mong Njua*), which also has a large number of speakers in Southeast Asia and diaspora communities. In China, the language is typically called *Miao*.

The language described below is representative of the White Hmong spoken by diaspora communities in the United States, in particular those residing in Minnesota. The Hmong community in the USA was founded in the late 1970s with a wave of emigration from Laos (often via Thailand). In his dictionary, Heimbach (1969: xi–xiv, xvii–xxiii) provides some basic information about the phonology of White Hmong; Ratliff (1992: 8–13) provides a different analysis, which we largely follow here. Descriptions of the phonological inventory can also be found in Smalley 1976, Lyman 1979, and Fulop & Golston 2008.

The most widespread Hmong orthography is the Romanized Popular Alphabet (RPA); other orthographies such as *Nyiakeng Puachue* Hmong and *Pahawh* Hmong are used almost

Table 1 White Hmong vowel inventory (monophthongs only).

	Front	Central	Back
Close	i	ɨ	u
Close-mid	e ẽ		õ
Open-mid			ɔ
Open		a	

Table 2 White Hmong tonal inventory.

Traditional description	Description (Esposito 2012)	Orthographic symbol	Example
High-level (55)	High-rising (45)	-b	*tob* [tɔ 45] 'deep'
Mid-level (33)	Mid-level (33)	ø	*to* [tɔ 33] 'pierced'
Low-level (22)	Low-level (22)	-s	*tos* [tɔ 22] 'wait'
High-falling (52)	High-falling (52)	-j	*toj* [tɔ 52] 'hill'
Mid-rising (24)	Mid-rising (24)	-v	*tov* [tɔ 24] 'mix'
Falling creaky (21)	Low-falling creaky (21)	-m	*tom* [tɔ̰ 21] 'bite'
Falling breathy (42)	High-falling breathy (52, 42)	-g	*tog* [tɔ̤ 42] 'sink'
Falling-rising (214)	–	-d	*tod* [tɔ 214] 'over there'

exclusively in Southeast Asia and China. Prior to the creation of the RPA in the early 1950s, the language was unwritten. More information regarding the orthographies, in particular *Pahawh* Hmong, can be found in Smalley, Vang & Yang 1990.

1.2 Vowel segmental inventory

White Hmong has six oral vowels /i e a ɔ u ɨ/ and two nasal vowels /ẽ õ/ (Table 1), as well as five diphthongs, three of which are falling /ai au aɨ/ and two rising /ia ua/. As is common in Hmong-Mien languages (e.g. Liu et al. 2020), as well as in other languages (e.g. Brazilian Portuguese, Barlaz et al. 2018), the nasal vowels are frequently accompanied by a velar nasal coda: /ẽ õ/ → [ẽ(ŋ) õ(ŋ)]. These nasal codas are treated as part of the vowel, because they are only found following the close-mid nasal vowels. The closely related variety Green Mong has a similar vowel inventory, with a couple of notable differences: it has an additional nasal vowel /ã/, which in White Hmong has merged with oral /a/ (compare White Hmong *paj* /pa 52/ and Green Mong *paaj* /pã 52/ 'flower'); moreover, Green Mong has merged both /ia a/ to /a/ (compare White Hmong *tshiab* /t͡ʂʰia 45/ and Green Mong *tshab* /t͡ʂʰa 45/ 'new').

1.3 Tonal inventory and tone sandhi

White Hmong morphemes are predictably (C)V(V) in structure. There are no phonemic coda consonants, though coda [ŋ] is attested as a phonetic implementation of the nasal vowels, as mentioned above. Onsets can have a very complex internal structure. (A complete list of White Hmong consonants also available as supplementary materials at https://pages. ucsd.edu/~mgarellek/files/SupplementaryFiles_Hmong.html.) Additionally, morphemes are obligatorily specified with one of eight tones, of which all but one are lexical; see Table 2.

All tones except the falling-rising -*d* (214) tone are lexical; the falling-rising -*d* (214) tone is not productive, but is found on certain high-frequency words (typically deictic markers). This tone developed from a merger of two morphemes bearing the low-falling creaky -*m* (21) tone followed by the mid-rising -*v* (24) tone (Ratliff 1992).

The falling tones with non-modal phonation appear to be consistently breathy or creaky (Esposito 2012, Garellek et al. 2013, Garellek, Keating & Esposito 2014). However, in earlier

work (e.g. Huffman 1987) the low-falling creaky -*m* (21) tone is called 'checked,' i.e. it is analyzed as consisting of a short but modal vowel followed by a glottal stop. Huffman based this analysis on phonetic grounds, but it is also worth mentioning that the low-falling creaky -*m* (21) tone is derived historically from syllables closed by a glottal stop (Ratliff 2010). Further support for this analysis may come from tone identification studies, which found that listeners do not rely on creaky voice to perceive the low-falling creaky -*m* (21) tone. Instead, listeners pay attention to f0 and duration (Garellek et al. 2013, 2014).

These tones are also found in Green Mong (Andruski & Ratliff 2000, Andruski & Costello 2004, Andruski 2006, Mai, Moua & Garellek 2019), which have similar pitch contours. The main differences are the following: in Green Mong, the falling breathy tone -*g* is mid-falling (31) instead of high-falling; moreover, non-modal phonation associated with tones (both breathiness and creakiness) is not consistently found in Green Mong (Andruski 2006).

Aside from their pitch and voice quality, White Hmong tones also differ in duration. Esposito (2012) found that the high-rising -*b* (45) and low-falling creaky -*m* (21) tones are on average shorter than the remaining lexical tones. (She did not measure the duration of the falling-rising -*d* tone.) Vowel quality differences as a function of tone are unattested in White Hmong, but have been known to occur diachronically in the distantly-related Hmong variety known as Shuijingping Mang (Mortensen 2013).

Hmong also possesses tone sandhi. However, as Ratliff (1987) points out, tonal alterations do not neatly conform to the definition of 'sandhi' proposed by Pike (1948). The main issue is that, unlike other cases of tone sandhi, the environment alone is not enough to trigger the tonal change; in White Hmong, other elements such as grammatical category and/or the particular lexical item also play a role. Despite these complexities, White Hmong tone sandhi can be summarized as follows: when the preceding tone is either high-rising -*b* (45) or high-falling modal -*j* (52), the following alterations occur: high-falling modal -*j* (52), low -*s* (22), and low-falling creaky -*m* (21) tones become high-falling breathy -*g* (42~52), whereas mid-rising -*v* (24) becomes mid-level Ø (henceforth denoted as -*x*, 33) and mid-level -*x* (33) becomes low -*s* (22). A detailed account of tone sandhi in White Hmong may be found in Ratliff 1992.

2 Methods

2.1 Speakers

Eight speakers of White Hmong were recorded. Subjects were recruited through personal contacts. Speakers ranged from 25 to 36 years of age and were born in Laos, Thailand or the US, but currently reside in Minneapolis/St. Paul, Minnesota. All speakers spoke English in addition to White Hmong; the reported age of English onset ranged from 5 to 16 years of age. All speakers reported that they used White Hmong daily. Speakers learned to read the Romanized Popular Alphabet (RPA) through (i) the Hmong Literacy and Culture Program, a Twin Cities public school program designed to give students the opportunity to learn aspects of the Hmong language and culture, including literacy, as part of the regular school day; (ii) at home through family instruction; or (iii) were self-taught, as is the case for Speaker 2. Table 3 below summarizes the background information on the speakers; self-reported gender, age, birthplace, number of years in the United States, and the reported age of English onset are given.

2.2 Corpus

The data in this study come from a corpus of read speech. The study's participants were recorded reading three folk tales that are freely available from the D.C. Everest Oral History Project (https://www.dceoralhistoryproject.org/, accessed 23 April 2016). These tales were

Table 3 Biographical information about the speakers in the corpus.

Speaker	Gender	Age (years, approx.)	Birthplace	Years in US	Age of English onset (years)
1	M	31	US	–	5
2	M	35	Laos	30	6
3	F	30	US	–	5
4	M	28	Laos	10	20
5	M	25	Thailand	10	16
6	F	35	US	–	5
7	F	36	Laos	20	16
8	F	32	Laos	27	5

chosen because they are written in a colloquial style. A native speaker verified that these were common stories, that were all known in the Hmong community. A copy of each story is shown in supplementary materials.

2.3 Procedure

All recordings were made in a sound-attenuated booth at Macalester College, St. Paul, Minnesota. Participants were asked to read the stories as naturally as possible, and were allowed to read over the story before the recording began.

After the recording was complete, the vowel of each word (N = 11,282 tokens) was manually segmented; for nasal vowels (N = 673 tokens), we included the nasal coda when present so as not to cut off the tonal contour. The onset of the vowel was taken to be the end of a transient release (for preceding stops), end of frication noise (for fricatives and aspirated stops), or onset of clear vowel formants (for preceding sonorants). The offset of the vowel was taken to be the end of clear vowel formants and/or onset of consonantal constriction.

Vowels were further coded for position in utterance: utterance-initial and utterance-final words were those adjacent to a visible breath. In total, there were 1607 initial tokens, 7814 medial ones, and 1766 final ones. There were an additional 95 words that counted as whole utterances, because they were preceded and followed by a breath. These were excluded from subsequent analyses of f0 and phrasing. The presence of an utterance-medial phrase boundary adjacent to the word was also coded, based on perceived phrase-final lengthening. In total, there were 1356 such tokens, which were excluded from any analyses of f0 and phrasing.

We also coded for the presence of a disfluency (defined by the presence of a prolongation, restart, or abrupt cutoff) on or adjacent to a target word; such tokens (N = 350, 3% of all tokens) were excluded from all subsequent analyses.

All vowels were subsequently annotated and analyzed acoustically using VoiceSauce (Shue et al. 2011). The acoustic measures used in the current student include: the frequencies of the first and second formants (F1, F2), calculated using Snack (Sjölander 2004); fundamental frequency (f0) as calculated by the STRAIGHT algorithm (Kawahara, de Cheveigné & Patterson 1998); H1*–H2*, the difference in amplitude between the first two harmonics corrected for the effects of surrounding formant frequencies and bandwidths (Hanson 1995) using the algorithm developed by Iseli, Shue & Alwan (2007); Cepstral Peak Prominence (CPP), a harmonics-to-noise ratio (Hillenbrand, Cleveland & Erickson 1994); and A1–P0, the difference in amplitude between the first formant and the first nasal pole (the harmonic closest to 250 Hz; Chen 1997). Fundamental frequency, formant frequencies, and H1*–H2* were calculated automatically at every millisecond of target vowel duration, and then averaged within thirds of the vowel of the vowel's duration. A1–P0 was calculated from VoiceSauce measures of H1*, H2*, and A1*, following the method described in Garellek, Ritchart & Kuang 2016.

Tokens with f0 outliers were excluded from the analysis of f0 and H1*–H2*. These include any token whose f0 was greater than absolute 2.5 standard deviations from a given speaker's mean f0. In total, 187 such outliers (less than 2% of all tokens) were identified and removed. Tokens with formant outliers were excluded from the vowel quality, nasality, and H1*–H2* analyses (Sections 3, 3.2, and 4.2 below). Outliers include those with mean F1 and F2 values greater than absolute 2.5 standard deviations from the grand mean, as well as those deemed outliers from visual inspection of the vowel plots. We excluded 2424 such outliers, or roughly 21% of all tokens.

3 Vowel quality

The first and second formants of monophthongal vowels of White Hmong are shown in Figure 1, and descriptive statistics are included in Table 4. The oral vowels are well dispersed from one another. The nasal vowel /ẽ/ has an overlapping distribution with both /e/ and /ɨ/ in the F1~F2 space. However, we show below that /ẽ/ is longer than both vowels, and is also heavily nasalized, usually frequently ending in a nasal coda [ẽ(ŋ)].
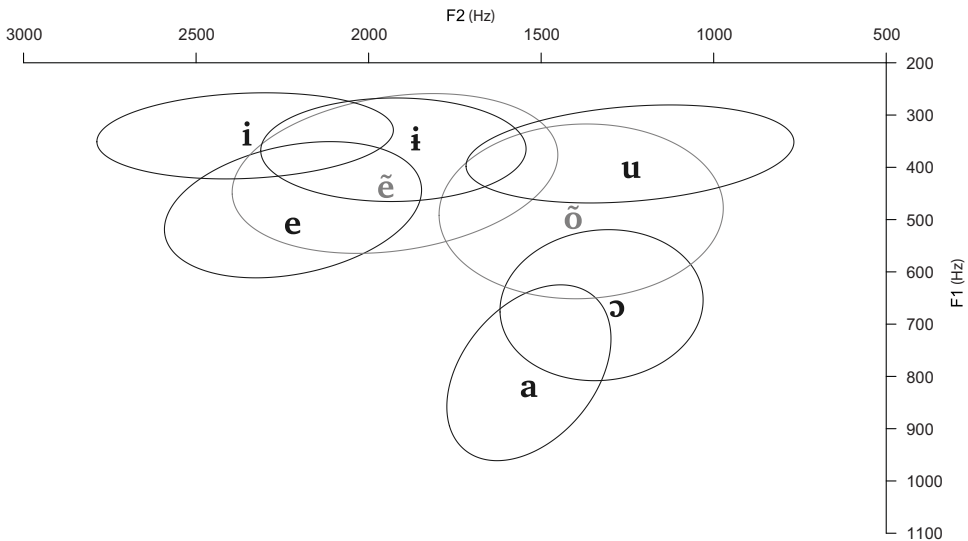

Figure 1 F1~F2 space in Hz for the oral and nasal monophthongal vowels. Ellipses represent 68% confidence intervals about the mean.

Table 4 Monophthong means and standard deviations (SD) of F1, F2, and duration.

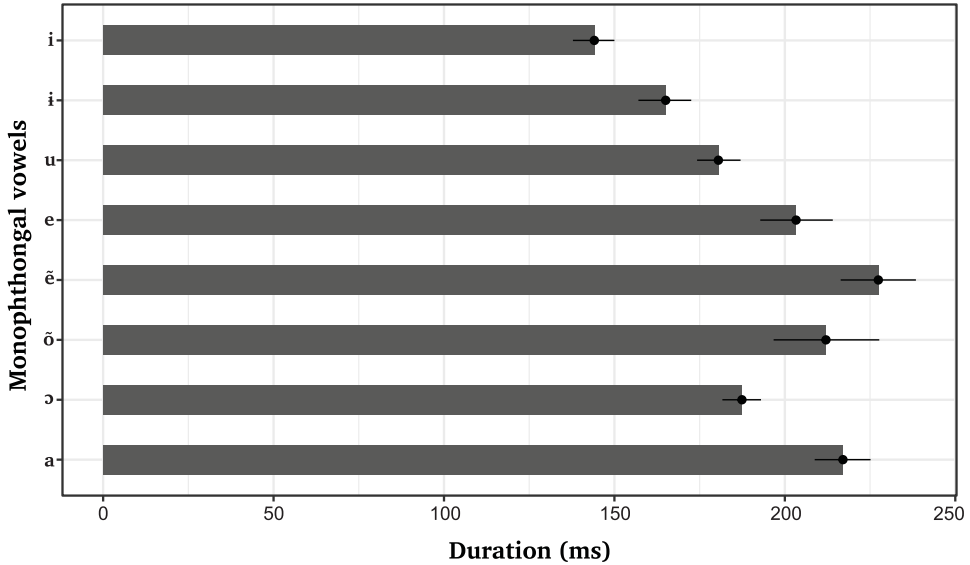| Vowel	Number of tokens analyzed	Mean F1 (SD) in Hz	Mean F2 (SD) in Hz	Mean duration (SD) in ms
/i/	965	340 (54)	2358 (283)	144 (92)
/e/	471	481 (86)	2219 (245)	203 (167)
/a/	661	793 (111)	1535 (157)	217 (105)
/ɔ/	1213	664 (95)	1325 (194)	187 (104)
/u/	1005	374 (62)	1243 (313)	180 (108)
/ɨ/	613	366 (65)	1928 (253)	165 (102)
/ẽ/	350	412 (101)	1924 (310)	227 (103)
/õ/	132	484 (109)	1384 (269)	212 (94)

Figure 2 Duration of monophthongal vowels. Error bars represent bootstrapped 95% confidence intervals about the mean.

The monophthongs also differ in duration, as illustrated in Figure 2. As expected (Peterson & Lehiste 1960, Umeda 1975), the high vowels /i ɨ u/ are generally shorter than mid oral vowels /e ɔ/, which in turn are shorter than /a/. However, nasal vowels /ẽ õ/ are longer than oral ones.

3.1 Diphthongs

The diphthongs of White Hmong are shown in Figure 3, overlaid atop the formant ellipses of the oral monophthongs (as shown in Figure 1). In total, we analyzed 184 tokens of /ai/, 688 tokens of /au/, 551 tokens of /aɨ/, 879 tokens of /ia/, and 1051 tokens of /ua/.

Diphthongs are generally more centralized in the vowel space than monophthongs. Although typically transcribed as /ia ua/ the rising diphthongs are acoustically realized more like [ĕə ŏɔ], respectively. Note that the 'extra-short' diacritic here is used to indicate which
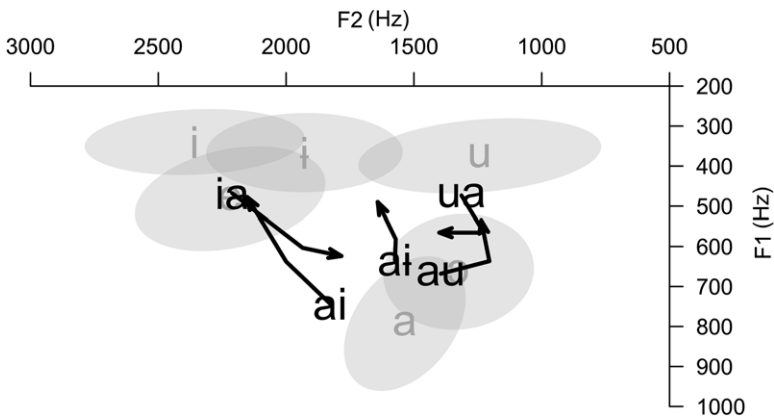

Figure 3 Trajectory of diphthongs over vowel thirds, overlaid on F1~F2 space in Hz for the oral monophthongal vowels. Ellipses represent 68% confidence intervals about the mean for monophthongs.

portion of the diphthong is shortest, based on trajectories in Figure 3. For example, /ia ua/ are characterized by relatively larger formant changes between the first and second thirds of the vowel, followed by relatively smaller changes in the last two thirds. This indicates that the quality of the first half of the diphthong is shorter than that of the second half.

The falling diphthongs, typically transcribed as /ai aɨ au/, are generally realized more like [ae ɐə ɔŏ], respectively. Given that /au/'s formant trajectory lies within the range of values for monophthongal /ɔ/, one could also describe the former vowel as [ɔː].

3.2 Nasality

Earlier in this section, we showed how the nasal vowels /ẽ õ/ are longer and more centralized than their oral counterparts. We also claimed that they are frequently realized with a velar nasal coda, as in /ẽ õ/ → [ẽ(ŋ) õ(ŋ)]. These nasal codas are treated as part of the vowel, because they are only found following the mid-close nasal vowels. However, it is clear that the nasal vowels are heavily nasalized throughout their duration. Figure 4 illustrates the time course of A1–P0 over the course of oral and nasal mid vowels. Lower values of the measure indicate greater acoustic nasality (Styler 2017). As seen in Figure 4, /ẽ õ/ both have lower values of A1–P0 relative to their oral counterparts /e ɔ/, even at the beginning of the vowel. This indicates that nasal vowels are nasalized throughout their duration. It is unclear why /ẽ/ has overall lower values of A1–P0 (i.e. is more acoustically nasalized) than /õ/; perhaps this serves to further distinguish /ẽ/ from both /e/ and /ɨ/, with which it overlaps substantially in terms of F1 and F2 (see Figure 1 above).
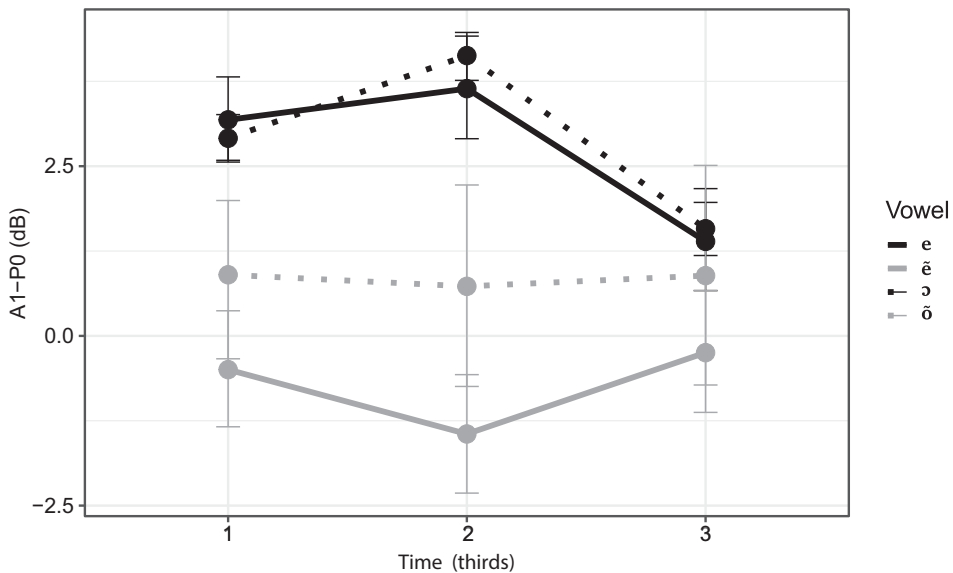

Figure 4 Mean A1–P0 (in dB) over the course of the first, middle, and final thirds of the vowel. Error bars represent bootstrapped 95% confidence intervals about the mean. Lower values indicate greater acoustic nasality.

4 Tone

4.1 Fundamental frequency

The f0 contours of the seven lexical tones in White Hmong are shown in Figure 5. Overall they conform well to the descriptions by Esposito (2012), who provided revised Chao numerals for the tones based on their citation forms.
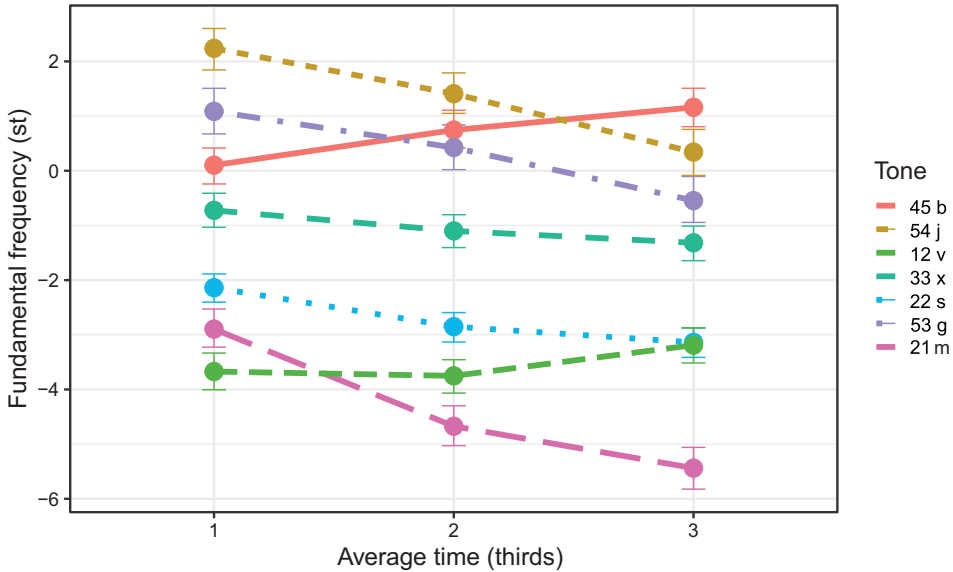
Figure 5 (Colour online) Mean f0 (in semitones, base = 200 Hz) over the course of the first, middle, and final thirds of the vowel by tone, represented using Chao numerals and their orthographic symbol. The mid-level tone is represented here with an '*x*' but is unmarked in the orthography. Error bars represent bootstrapped 95% confidence intervals about the mean.

We highlight here some differences with the contours in our corpus and those represented in earlier work (e.g. Ratliff 1992, Esposito 2012). The high(-rising) -*b* tone is clearly high-rising, and thus is best represented in Chao numerals as 45, as in Esposito 2012. The high-falling breathy -*g* tone falls to the same f0 as the mid tone, and thus should be represented as 53 (rather than 52). The f0 of the high-falling modal -*j* tone, typically described as 52, is scaled slightly higher than that of the high-falling breathy tone; it should probably be represented as 54. Finally, the mid-rising -*v* tone, typically described as 24, should probably be labeled as a low-rising 12, as it rises to the same f0 height as the low-level (22) -*s* tone. Our proposed revision to the tone numerals in White Hmong, based not just on these contours but also on phrasal variants discussed in Section 4.1.2 below, is shown in Table 6 in Section 5.2. In the figures shown here, tone labels correspond to our proposed revisions.

The eighth tone, which is used in a closed class of deictic markers such as *tod* 'over there', is traditionally described as having a low-falling-rising 214 contour. Indeed, the tone is derived from blending of two words, one with the low-falling creaky (21) -*m* tone, and the other with a low-rising (12, traditionally transcribed as 24) -*v* tone (Ratliff 1992). In our dataset, this tone is clearly low-rising, and has a similar contour to the low-rising tone. But as shown in Figure 6, the -*d* tone rises to a slightly higher f0 than the open-class low-rising tone. For this reason, we transcribe it as having a 13 contour, instead of the 214 contour. This accords with the Chao letters used in earlier tonal descriptions of the language (Heimbach 1969).

Next we describe how the lexical tones vary as a function of position in utterance, as seen in Figure 7. (We exclude tokens with the -*d* tone because of its limited distribution in the dataset.) Several generalizations emerge from the tonal contours shown in that figure. First, all tones are scaled higher in utterance-medial position compared to tones in utterance-initial and utterance-final positions; see also Table 5. Tones in utterance-initial vs. utterance-final position do not differ in scaling, except for the low-level -*s* (22) and low-falling -*m* (21) tones, which are scaled lower in utterance-final position compared to in utterance-initial position. Given the cross-linguistic tendency for f0 to decline over the course of an utterance
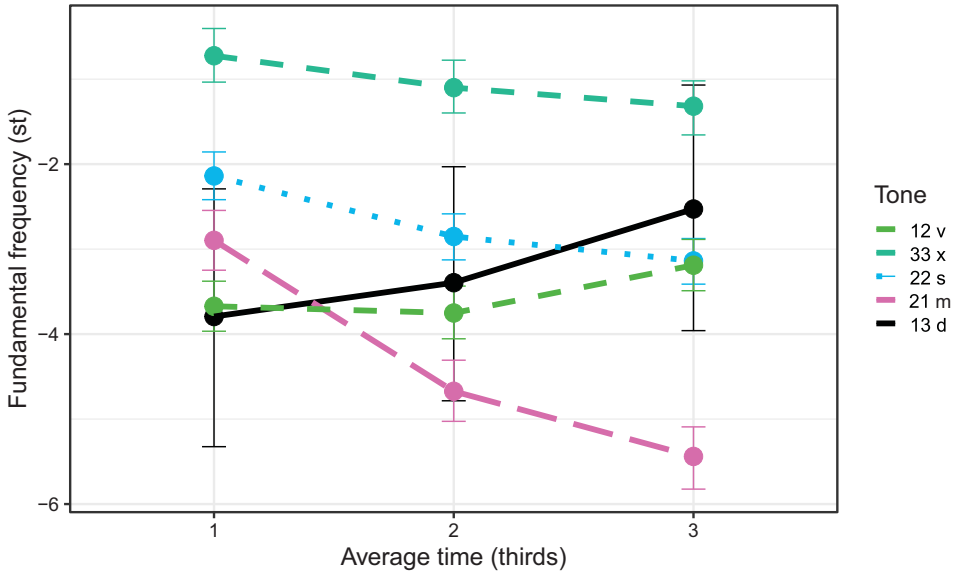
Figure 6 (Colour online) Mean f0 (in semitones, base = 200 Hz) over the course of the first, middle, and final thirds of the vowel for the non-productive *-d* tone and other low-register tones, represented using Chao numerals and their orthographic symbols. The mid-level tone is represented here with an '*x*' but is unmarked in the orthography. Error bars represent bootstrapped 95% confidence intervals about the mean.

(Ladd 1988, Yuan & Liberman 2014), the data suggest that White Hmong tones are scaled lower than expected in utterance-initial position; if the variation in pitch scaling were due to declination, then we would expect tones in utterance-initial position to have the highest pitch range.

Another generalization concerns the similarity in pitch contour between the two high-falling tones (modal *-j* vs. breathy *-g*), and their scaling relative to the high-rising *-b* tone. The two high-falling tones are often described as having essentially the same contour, either 42 or 52 (Ratliff 1992, Esposito 2012, Garellek et al. 2013). In utterance-initial and utterance-final positions, their contours do indeed overlap (see f0 values in Table 5). In utterance-medial position, however, the modal high-falling *-j* tone is scaled higher than the breathy *-g* tone. Thus, the two high-falling tones often overlap in pitch contour, but this varies as a function of phrasing.

Further, these tones are not consistently 52 in contour, as described in previous work. In Figure 7 we see that these tones end on a frequency similar to that of the mid-level tone, but only in utterance-final position (see also Table 5). In other phrasal positions, these tones generally end on a higher level, warranting a 54 or upstepped ^54 label; that is, the high-falling tones can end at the same high-f0 level as the high-rising *-b* (45) tone.

Another interesting finding to emerge in the data is that, in utterance-initial position, the mid *-s* (33) tone and the low *-s* (22) tone have similar f0 levels, with both corresponding to a mid-level tone (see Table 5). Thus, if these tones are distinguishable to listeners in utterance-initial position, this begs the question of what cues they are attending to.

Finally, it is worth mentioning that, in utterance-final position, the tonal contours are qualitatively very similar to those found in utterance-medial or utterance-initial position. This suggests that the ends of utterances are not marked intonationally via the addition of boundary tones, at least for the types of sentences found in these stories.
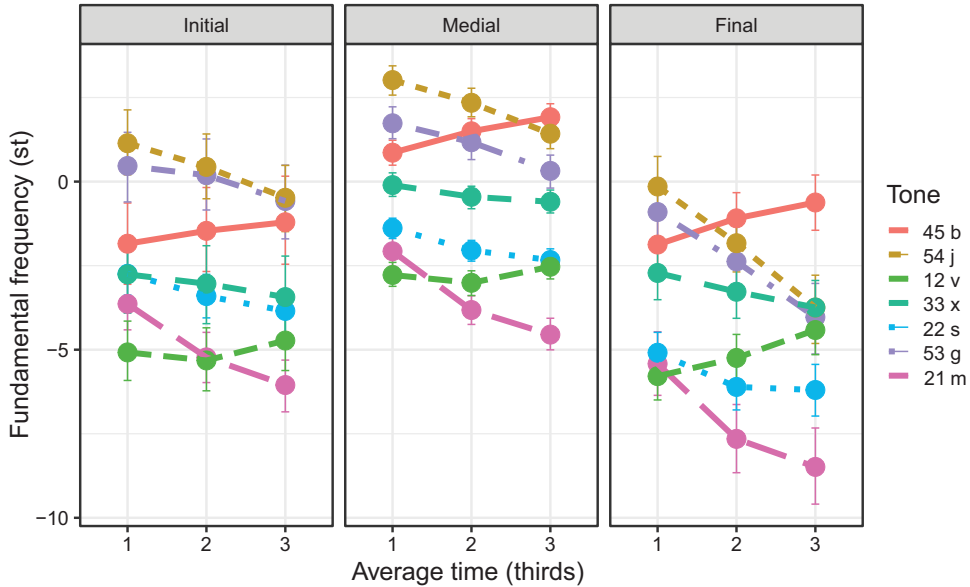
Figure 7 (Colour online) Mean f0 (in semitones, base = 200 Hz) over the course of the vowel by tone and position in utterance (Initial, Medial, Final). The mid-level tone is represented here with an '*x*' but is unmarked in the orthography. Error bars represent bootstrapped 95% confidence intervals about the mean.

Table 5 Means (first third-last third) and overall standard deviations (SD) of tone's f0, according to utterance position.

Tone	Total number of tokens	Utterance-initial mean f0 (SD) in Hz	Utterance-medial mean f0 (SD) in Hz	Utterance-final mean f0 (SD) in Hz
High rising -*b*	1518	195–205 (83)	224–238 (74)	193–209 (76)
High-falling modal -*j*	1216	230–211 (85)	253–232 (77)	212–174 (70)
Low-rising -*v*	1646	161–165 (63)	181–183 (57)	154–167 (58)
Mid -*x*	1600	184–177 (68)	211–205 (64)	183–174 (62)
Low -*s*	2208	183–173 (66)	196–186 (59)	159–151 (57)
High-falling breathy -*g*	1077	223–211 (85)	236–218 (76)	207–174 (78)
Low-falling creaky -*m*	1313	175–153 (64)	189–165 (58)	157–137 (59)

4.2 Phonation

In this section, we seek to describe, in more detail than previously reported, the non-modal phonation associated with the high-falling breathy -*g* (53) tone and the low-falling creaky -*m* (21) tone in White Hmong. Earlier work has shown for citations tones that the high-falling breathy -*g* tone is consistently breathy (relative to the high-falling modal -*j* tone), and that the low-falling creaky -*m* (21) tone is consistently creaky relative to the low modal -*s* (22) tone (Esposito 2012, Garellek 2012; see also Huffman (1987), who found that the low-falling tone was not consistently creaky, and Andruski (2006), who found that some speakers of Green Mong produce this tone with modal voice). Moreover, Garellek et al. (2013, 2014) showed that listeners rely on the presence of breathiness to identify the high-falling breathy -*g* tone. However, they ignore the presence of creaky voice when identifying the low-falling creaky -*m* tone, relying instead on low f0 and short vowel duration. Thus, we would expect that, for non-citation forms, the high-falling breathy -*g* tone should be consistently breathy in all utterance positions. On the other hand, the low-falling creaky -*m* tone might not be consistently creaky in non-citation form, because listeners do not rely on breathy voice when identifying it.
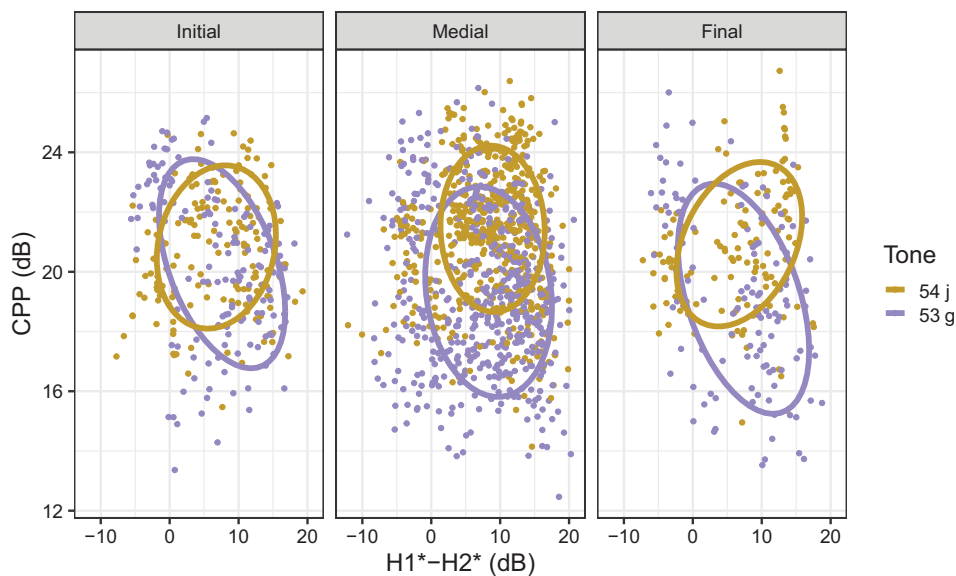
Figure 8 (Colour online) H1*–H2* (x-axis) by CPP (y-axis) measurements for the two high-falling tones. The bottom-right of each facet represents the breathiest acoustic quality. Ellipses represent 68% confidence intervals about the mean.

In the previous sections, we saw that most f0 variation in White Hmong tones tends to be driven by phrasing. Thus, here we describe voice quality differences as a function of position in utterance. In Figures 8 and 9, we show a two-dimensional acoustic space of the voice quality, with CPP on the y-axis and H1*–H2* on the x-axis. CPP measures spectral noise, which can be derived from aspiration noise during breathiness, and from irregular voicing during creakiness (Garellek 2019). To our knowledge, CPP is not sensitive to differences (e.g. from aspiration vs. irregular f0) in the noise spectrum. Representing both spectral noise (as measured by CPP) and low-frequency spectral tilt (as measured by H1*–H2*) in the same figure provides a more comprehensive interpretation of the acoustics of voice quality than if only one is looked at in isolation (Garellek 2019, Seyfarth & Garellek 2018). In Hmong, the high-falling breathy (53) -*g* tone never sounds creaky, and low-falling creaky (-*m*) tone never sounds breathy. Consequently, the lower CPP expected for the high-falling breathy tone (relative to its modal counterpart) will be interpreted as acoustic evidence for BREATHINESS; conversely, the lower CPP expected for the low-falling creaky tone (relative to its modal counterpart) will be interpreted as acoustic evidence for CREAKINESS.

H1*–H2* requires accurate estimation of f0 (for determining the harmonics) as well as formants (for the correction, see Section 2). Thus, we excluded tokens whose f0, F1, and F2 values exceeded 2.5 standard deviations from the mean for these three measures. For the following analysis we also excluded nasal vowels (whose acoustics can interact with breathiness; Chen 1997, Garellek et al. 2016, Styler 2017).

In Figure 8, we compare the high-falling breathy (53) -*g* tone and its modal (54) -*j* tone counterpart. The breathiest part of each plot is the bottom-right corner, where H1*–H2* is highest (higher spectral tilt, correlated with increased vocal fold spreading) and where CPP is lowest (higher spectral noise, in the case of breathy vowels correlated with increased aspiration). In initial and final positions, we see that there is quite a bit of overlap between the two high-falling tones. However, in the bottom-right of the figure, there are generally more tokens of the breathy -*g* tone than the modal -*j* tone. In medial position, the voice quality difference between these two tones appears to be driven largely by the lower CPP for the breathy -*g* tone, rather than by a combination of CPP and H1*–H2*. Thus, the breathy -*g* tone
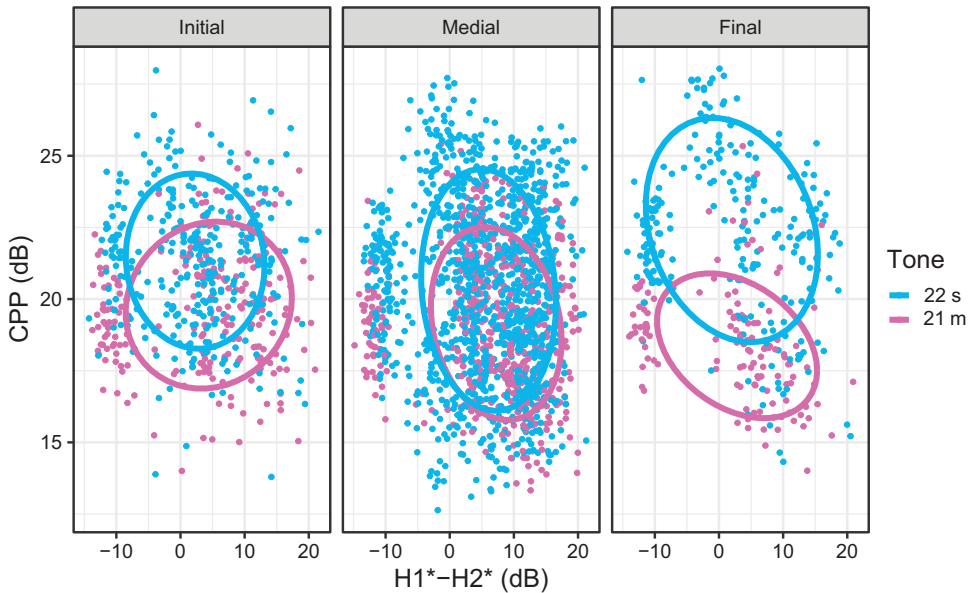
Figure 9 (Colour online) H1*–H2* (x-axis) by CPP (y-axis) measurements for the two low tones. The bottom-left of each facet represents the creakiest acoustic quality. Ellipses represent 68% confidence intervals about the mean.

is consistently noisier than its modal counterpart across phrasal positions; however, it is not consistently produced with higher spectral tilt.

A similar picture comparing the voice quality of the low-falling creaky (-*m*) tone vs. the low-level modal (-*s*) tone is seen in Figure 9. The creakiest part of each plot is the bottom-left corner, where H1*–H2* is lowest (lower spectral tilt, correlated with increased vocal fold constriction) and where CPP is lowest (higher spectral noise, correlated with increased voicing irregularity in creaky vowels). In utterance-final position, the low-falling creaky tone is generally lower in CPP, as expected given the increased noise due to irregular voicing. However, the two tones overlap greatly in terms of H1*–H2*. In other phrasal positions, the same general pattern is found, only the overlap in terms of both H1*–H2* and CPP is greater.

In sum, the two tones with non-modal phonation show more gradient, rather than categorical, differences in voice quality from their modal counterparts, particularly in non-final positions. Breathiness does not appear to be more robust to phrasal position than creakiness. Voice quality differences are more pronounced acoustically in terms of spectral noise (here measured by CPP) than by spectral tilt, as measured by H1*–H2*. This may be due to the fact that H1*–H2* is particularly sensitive to accurate estimation of f0, formant frequencies, and formant bandwidths, which can lead to error propagation (Chai & Garellek 2019). We discuss this more in Section 5.

4.3 Duration

In this corpus of read speech, average tone duration interacts with vowel quality, see Figure 10. In previous work (Huffman 1987, Ratliff 1992, Esposito 2012), it was shown that, in citation forms, the high-rising -*b* (45) tone and the low-falling creaky -*m* (21) tone were on average shorter than other lexical tones. In our data however, main effects of tone cannot be found, as tonal duration clearly interacts with vowel quality. For instance, while the high-rising -*b* (45) tone and the low-falling creaky -*m* (21) are indeed shorter than other tones on /i/, those same tones tend to be longer on /u/.
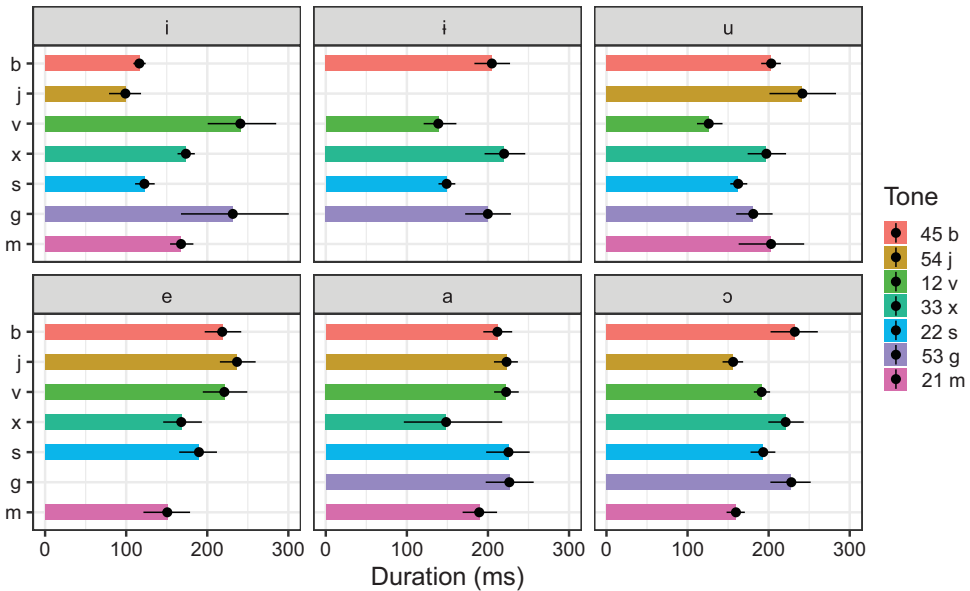
Figure 10 (Colour online) Duration of tones in utterance-medial position, grouped by oral monophthongal vowels. Error bars represent bootstrapped 95% confidence intervals about the mean.

Ratliff also reported a difference between the low-level *-s* (22) and mid-level (33) tone: 'The difference between the -s tone (22) and the -Ø tone (33) seems to involve something other than pitch. It may be duration: there is a certain "chanted" quality to the mid level tone, which has no perceivable fall at the end, while there is a more natural tapering at the end of the low level tone' (Ratliff 1992: 11). We suspect this is due to the f0 contour, rather than duration; Figure 10 shows that, for some vowels, the low-level *-s* (22) tone is longer than the mid-level *-x* (33). However, both tones tend to have a slight fall; see Figure 7 above.

However, we caution against coming to firm conclusions regarding tone duration from this corpus, in which tones are presumably unbalanced for lexical frequency and part of speech, among other factors that are well known to influence duration.

5 Discussion and conclusions

The primary goal of this study was to advance the quantitative description of the vocalic and suprasegmental features of White Hmong. Although tone and phonation in citation forms are well described for the language (Huffman 1987; Ratliff 1992; Esposito 2012; Garellek et al. 2013, 2014), this is the first study to quantify and describe these features in a broader context. Our results reveal novel findings of descriptive and theoretical relevance for all features investigated.

5.1 Vowel nasalization and quality

We find evidence that nasal vowels are indeed acoustically nasalized throughout their duration (Figure 4). Given that nasal vowels often end in a nasal coda (e.g. /ẽ/ → [ẽŋ]), there has been some discussion in the literature as to whether the nasal coda should be treated as phonemic or not in both White Hmong and Green Mong (Andruski & Ratliff 2000; see similar discussion for nasal vowels in Brazilian Portugese; Barlaz et al. 2018, Marques & Scarborough 2020). However, sources generally treat these vowels as phonemic nasal vowels

Table 6 Proposed revisions to White Hmong tone descriptions.

Traditional description	Description (Esposito 2012)	Proposed revisions	Orthographic symbol
High-level (55)	High-rising (45)	Always high-rising (45)	*-b*
Mid-level (33)	Mid-level (33)	Mid-level (33) Low-level (22) utterance-initially	ø (denoted as *-x*)
Low-level (22)	Low-level (22)	Low-level (22)	*-s*
High-falling (52)	High-falling (52)	High-falling (54∼53) Upstepped high-falling (^54) utterance-medially	*-j*
Mid-rising (24)	Mid-rising (24)	Low-rising (12∼23)	*-v*
Falling creaky (21)	Low-falling creaky (21)	Low-falling creaky (21)	*-m*
Falling breathy (42)	High-falling breathy (52, 42)	High-falling breathy (54∼53)	*-g*
Falling-rising (214)		Low-rising (13)	*-d*

with an optional nasal coda, largely based on phonotactic grounds (Andruski & Ratliff 2000). Our study supports this analysis and confirms that the vowels are nasalized throughout their duration. Nasal vowels also tend to be longer and more centralized than oral ones.

Our data reveal that diphthongs are quite centralized relative to oral monophthongs, both in terms of the diphthongs' nucleus and off-glide. In particular, the trajectory of the /au/ diphthong is firmly within the acoustic space of monophthongal /ɔ/, suggesting that this vowel could be transcribed phonetically as [ɔǒ∼ɔː] (Figure 3).

5.2 F0 as a function of tone and phrasing

The results on f0 realization reveal interesting properties about tones in White Hmong. First, there is little evidence for additional intonational boundary tones or other post-lexical f0 targets. Similar claims have been made for other languages with dense tonal systems such as Yoloxóchitl Mixtec (DiCanio, Benn & Castillo García 2018) and Itunyoso Triqui (DiCanio & Hatcher 2018); still, boundary tones are attested, in other languages with dense tone systems, e.g. Cantonese (Wai, Chan & Beckman 2005). Also, given that these stories were mostly composed of declarative and interrogative sentences, it is possible that extra intonational f0 targets are found in other sentence types. The few cases of exclamative sentences sometimes showed atypical f0 patterns, but they were too few to analyze systematically. More research on intonational tone targets (whether they occur, how they are realized, and how they interact with lexical tone) in White Hmong is therefore needed.

Still, there are clear effects of phrasal position within the utterance on tone realization (Figure 7). Utterance-initially, tones tend to be scaled LOWER than in medial position, which seems to go against patterns of declination, whereby the onsets of utterances have the highest pitch (Ladd 1988). We also find suggestions of f0 contour merger between the low-level *-s* (22) and mid-level *-x* (33) tones in utterance-initial position. Future work should therefore investigate how distinct these two tones are in this environment.

An interesting finding of this study is that the tone contours in citation forms sometimes differ quite largely from the contours measured here in read speech. As a result, we provide potential revisions to the descriptions of White Hmong tones, shown in Table 6. Some of these revisions are due to interactions with phrasing: for example, the mid-level *-x* tone is noted as 33 in citation forms (Esposito 2012), but we find it to be a low-level (22) tone in utterance-initial position. However, the two high-falling tones – modal *-j* and breathy *-g* – never end on

a low target, as reported for citation forms. Instead, they tend to be 54 or 53. Interestingly for the high-falling modal -*j* tone, we have evidence for the need to posit a sixth tonal level: in utterance-medial position, this tone starts higher than the end pitch of the high-rising -*b* (45) tone. So if that tone is labeled 45, the high-falling modal -*j* tone might be labeled as '65'. Here we settle on calling it an upstepped ^54, since we do not perceive falsetto voice commonly found on tones described as being level 6 in other languages (Zhu 2012): although tonal descriptions typically avoid positing more than 5 levels, 'ultra-high' tones produced with falsetto voice are attested in languages of East and Southeast Asia (e.g. in Hubei Chinese: Wang & Tang 2012; see also discussion in Zhu 2012).

5.3   Phonation type associated with the breathy and creaky tones

Previous studies of White Hmong phonation associated with two tones have shown that the high-falling breathy -*g* tone is consistently breathier than the high-falling modal -*j* tone, and that the low-falling creaky -*m* tone is consistently creakier the low modal -*s* tone (Esposito 2012, Garellek 2012; see also Huffman 1987, who, as noted previously, found that the low-falling tone was not consistently creaky). And in a series of perceptual studies, Garellek et al. (2013, 2014) showed that listeners rely on the presence of breathiness to identify the high-falling breathy -*g* tone, but that they ignore the presence of creak when identifying the low-falling creaky -*m* tone. Instead, listeners rely on low f0 and short vowel duration to perceive the low-falling tone. Based on these findings, we expected that the low-falling creaky -*m* tone would not be consistently creaky in read speech.

Contrary to this expectation, we find consistent non-modal phonation across phrasal positions for both the high-falling breathy -*g* and low-falling creaky -*m* tones (Figures 8 and 9). Despite the general tendency for the breathy tone to sound breathy and the creaky tone to sound creaky, ACOUSTIC differences in voice quality, as measured by spectral tilt and noise, are clearly gradient. For example, many modal tokens have more non-modal-like phonation than non-modal ones, presumably because voice acoustics (and production) varies on a continuum from more modal-like to more non-modal-like (Garellek 2019, Esposito & Khan 2020). The acoustic difference in voice quality between modal vs. non-modal tones is also most pronounced utterance-finally. That listeners ignore creakiness in favor of other tonal cues (namely pitch and duration), despite the fact that the creaky voice is robustly attested on the low-falling tone, might be due to this gradient differentiation between modal vs. non-modal voice quality. And as described in earlier work (e.g. Garellek et al. 2013), we suspect that listeners ignore creaky voice because creak is not perceptually distinct from low f0. Similar results are found in Cantonese, where listeners favor f0 cues over creak in the perception of the low and creaky tone (Zhang & Kirby 2020; see also Yu & Lam 2014).

Another notable finding of this study was that H1*–H2*, perhaps the most commonly used measure of non-modal phonation, was not effective at distinguishing modal vs. non-modal tones. In contrast, CPP did effectively distinguish modal vs. non-modal tones. Why might H1*–H2* be less effective than a noise measure? One possibility might be due to the calculation of the measures themselves. Unlike CPP, H1*–H2* requires adequate estimation of both f0 (for H1 and H2 calculation) as well as formant frequencies and bandwidths (see discussion in Garellek 2019, as well as in Chai & Garellek 2019). It also covaries with f0, but somewhat unpredictably in the lowest f0 range (Kuang 2017). In data sets with more variable speech then, spectral tilt measures like H1*–H2* (as well as other higher-frequency tilt measures) might be too noisy to provide clear generalizations. Owing to potential issues with the formant correction, one might think it preferable to use instead UNCORRECTED H1–H2, while controlling for vowel category. We caution against doing so. Although uncorrected H1–H2 avoids issues with formant correction, it re-introduces another source of unwanted variation that the correction was specifically designed to filter out: namely, variation in harmonic amplitude due to variation in formant frequencies and bandwidths. Models of uncorrected

H1–H2 that include vowel category as a control variable do not adequately control for effects of formant frequencies and their bandwidths.

It is also possible that the greater importance of spectral noise (over spectral tilt) to the phonation differences is a result of articulation-to-acoustics mapping. Let us assume that speakers target aspiration (along with a high-falling pitch) when producing the breathy tone, and low f0 in the case of the creaky tone (as suggested by Garellek et al. 2013). Aspiration is generally produced via increased vocal fold spreading, but H1–H2 (measured at the source) is affected more by changes to medial vocal fold thickness than by changes in glottal area (Zhang 2016). In the case of the creaky tone, low f0 has somewhat variable behavior on H1*–H2* (Kuang 2017), and can be produced by lower subglottal pressure, lower laryngeal height, as well as decreased anterior–posterior vocal fold stiffness, among other physiological mechanisms (Zhang 2016). Therefore, if Hmong speakers specifically target aspiration and low f0 when they produce the breathy and creaky tones, respectively, then these can be achieved through other means beyond changes in medial vocal fold thickness, which appear to be most important role in causing changes to H1–H2 and spectral tilt more generally.

5.4 Concluding remarks

This study provided an in-depth description of the vowels and tones in White Hmong, as they occur in whole utterances from read speech. We find convergences with previous descriptions (analyzing the nasal vowels as fully nasalized, despite the presence of a nasal coda), as well as noteworthy differences (the centralized quality of most diphthongs, the numeric description of the high-falling and rising tones), which perhaps derive from data collected from carrier phrases in previous studies compared to read speech in the current one.

In addition to furthering description of the vowel and tone systems of White Hmong, this study provides data that can be used to advance theoretical and methodological issues in phonetics more broadly. Clearly, there is the need for phonetic studies to include distinct speech styles (e.g. citation forms, isolated sentences, read speech of whole passages, and spontaneous speech) so as to provide more ecologically-valid descriptions of the sounds of a language. We also point to the need for incorporating phrasing in an analysis of a tone language's f0 contours; for considering an ultra-high tone level above level 5; for using noise over spectral tilt measures to analyze voice quality and phonation in less controlled data sets.

Acknowledgements

References

Andruski, Jean. 2006. Tone clarity in mixed pitch/phonation-type tones. *Journal of Phonetics* 34, 388–404.

Andruski, Jean & James Costello. 2004. Using polynomial equations to model pitch contour shape in lexical tones: An example from Green Mong. *Journal of the International Phonetic Association* 34(2), 125–140.

Andruski, Jean & Martha Ratliff. 2000. Phonation types in production of phonological tone: The case of Green Mong. *Journal of the International Phonetic Association* 30(1/2), 37–61.

Baran, Jane, Marsha Laufer & Ray Daniloff. 1977. Phonological contrastivity in conversation: A comparative study of voice onset time. *Journal of Phonetics* 5, 339–350.

Barlaz, Marissa, Ryan Shosted, Maojing Fu & Brad Sutton. 2018. Oropharyngeal articulation of phonemic and phonetic nasalization in Brazilian Portuguese. *Journal of Phonetics* 71, 81–97.

Brunelle, Marc. 2009. Northern and Southern Vietnamese tone coarticulation: A comparative case study. *Journal of Southeast Asian Linguistics* 1, 49–62.

Byrd, Dani. 1993. 54,000 American stops. *UCLA Working Papers in Phonetics* 83, 97–116.

Carroll, Lucien. 2015. *Ixpantepec Nieves Mixtec word prosody*. Ph.D. dissertation, University of California, San Diego.

Chai, Yuan & Marc Garellek. 2019. Using H1 instead of H1–H2 as an acoustic correlate of glottal constriction. Poster presented at the Fall 2019 Meeting of the Acoustical Society of America, San Diego, CA.

Chang, Yueh-chin & Feng-fan Hsieh. 2012. Tonal coarticulation in Malaysian Hokkien: A typological anomaly. *The Linguistic Review* 29(1), 37–73.

Chen, Marilyn. 1997. Acoustic correlates of nasalized vowels. *The Journal of the Acoustical Society of America* 102(4), 2360–2370.

Creswell, Cassandre & Kieran Snyder. 2000. Passive and passive-like constructions in Hmong. In Roger Billerey & Brook Danielle Lillehaugen (eds.), *Proceedings of the 19th West Coast Conference on Formal Linguistics* (WCCFL 19), 71–82. Somerville, MA: Cascadilla Press.

Crystal, Thomas & Arthur House. 1988. Segmental durations in connected-speech signals: Current results. *The Journal of the Acoustical Society of America* 83(4), 1553–1573.

Davidson, Lisa & Daniel Erker. 2014. Hiatus resolution in American English: The case against glide insertion. *Language* 90(2), 482–514.

DiCanio, Christian. 2014. Cue weight in the perception of Trique glottal consonants. *The Journal of the Acoustical Society of America* 119, 3059–3071.

DiCanio, Christian, Joshua Benn & Rey Castillo García. 2018. The phonetics of information structure in Yoloxóchitl Mixtec. *Journal of Phonetics* 68, 50–68.

DiCanio, Christian & Richard Hatcher. 2018. Is intonation universal? Colloquium talk presented at University of Massachussetts Amherst.

DiCanio, Christian T., Caicai Zhang, Douglas H. Whalen & Rey Castillo García. 2019. Phonetic structure in Yoloxóchitl Mixtec consonants. *Journal of the International Phonetic Association* 50, 333–365.

Dilley, Laura., Stephanie Shattuck-Hufnagel & Mari Ostendorf. 1996. Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics* 24, 423–444.

Esposito, Christina M. 2010. Variation in contrastive phonation in Santa Ana Del Valle Zapotec. *Journal of the International Phonetic Association* 40, 181–198.

Esposito, Christina M. 2012. An acoustic and electroglottographic study of White Hmong tone and phonation. *Journal of Phonetics* 40, 466–476.

Esposito, Christina M. & Sameer ud Dowla Khan. 2020. The cross-linguistic patterns of phonation types. *Language and Linguistics Compass* 14, e12392.

Fuller, Judith Wheaton. 1987. Topic markers in Hmong. *Linguistics of the Tibeto-Burman Area* 10(2), 113–127.

Fulop, Sean & Chris Golston. 2008. Breathy and whispery voicing in White Hmong. *Proceedings of Meetings on Acoustics* 4, 1–10.

Gandour, Jackson, Siripong Potisuk & Sumalee Dechongkit. 1994. Tonal coarticulation in Thai. *Journal of Phonetics* 22, 477–492.

Garellek, Marc. 2012. The timing and sequencing of coarticulated non-modal phonation in English and White Hmong. *Journal of Phonetics* 40, 152–161.

Garellek, Marc. 2014. Voice quality strengthening and glottalization. *Journal of Phonetics* 45(1), 106–113.

Garellek, Marc. 2019. The phonetics of voice. In William F. Katz & Peter F. Assmann (eds.), *The Routledge handbook of phonetics*, 75–106. Oxford & New York: Routledge.

Garellek, Marc, Patricia Keating & Christina M. Esposito. 2014. Relative role of phonation in White Hmong tone perception. In Oana David, Kayla Carpenter, Florian Lionnet, Christine Sheil, Tammy Stark & Vivian Wauters (eds.), *Proceedings of the 38th Annual Meeting of the Berkeley Linguistics Society* (BLS 38), 179–189. Berkeley, CA: Berkeley Linguistics Society.

Garellek, Marc, Patricia Keating, Christina Esposito & Jody Kreiman. 2013. Voice quality and tone identification in White Hmong. *The Journal of the Acoustical Society of America* 133, 1078–1089.

Garellek, Marc, Amanda Ritchart & Jianjing Kuang. 2016. Breathy voice during nasality: A cross-linguistic study. *Journal of Phonetics* 59, 110–121.

Hanson, Helen. 1995. *Glottal characteristics of female speakers*. Ph.D. dissertation, Harvard University.

Heimbach, Ernest E. 1969. *White Hmong–English dictionary* (Southeast Asia Program Publications). Ithaca, NY: Cornell University Press.

Hillenbrand, James M., Ronald A. Cleveland & Robert L. Erickson. 1994. Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research* 37, 769–778.

Huffman, Marie. 1987. Measures of phonation type in Hmong. *The Journal of the Acoustical Society of America* 81, 495–504.

Iseli, Markus, Yen-Liang Shue & Abeer Alwan. 2007. Age, sex, and vowel dependencies of acoustic measures related to the voice source. *The Journal of the Acoustical Society of America* 121, 2283–2295.

Jarkey, Nerida. 2010. Cotemporal serial verb constructions in White Hmong. In Mengistu Amberber, Brett Baker & Mark Harvey (eds.), *Complex predicates: Cross-linguistic perspectives on event structure*, 110–134. Cambridge: Cambridge University Press.

Kawahara, Hideki, Alain de Cheveigné & Roy Patterson. 1998. An instantaneous-frequency-based pitch extraction method for high-quality speech transformation: Revised TEMPO in the STRAIGHT suite. *5th International Conference on Spoken Language Processing* (ICSLP '98), Sydney, 0659.

Khan, Sameer ud Dowla. 2012. The phonetics of contrastive phonation in Gujarati. *Journal of Phonetics* 40, 780–795.

Kuang, Jianjing. 2017. Covariation between voice quality and pitch: revisiting the case of Mandarin creaky voice. *The Journal of the Acoustical Society of America* 142(3), 1693–1706

Ladd, D. Robert. 1988. Declination "reset" and the hierarchical organization of utterances. *The Journal of the Acoustical Society of America* 84, 530–544.

Lisker, Leigh & Arthur Abramson. 1967. Some effects of context on voice onset time in English stops. *Language and Speech* 10, 1–28.

Liu, Wen, You-Jing Lin, Zhenghui Yang & Jiangping Kong. 2020. Hmu (Xinzhai variety). *Journal of the International Phonetic Association* 50, 240–257.

Lyman, Thomas A. 1979. *Grammar of Mong Njua (Green Miao): A descriptive linguistic study*. Sattley, CA: Blue Oak Press.

Mai, Anna, Mai Moua & Marc Garellek. 2019. Tone-tune association and voice quality in Green Mong. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (ICPhS XIX), Melbourne, 1962–1966.

Marques, Luciana & Rebecca Scarborough. 2020. Perceptual compensation of vowel nasality in Brazilian Portuguese. *Journal of Portuguese Linguistics* 19(3), 1–18.

Mortensen, David. 2013. Tonally conditioned vowel raising in Shuijingping Mang. *Journal of East Asian Linguistics* 22(3), 189–216.

Peng, Shu-hui. 1997. Production and perception of Taiwanese tones in different tonal and prosodic contexts. *Journal of Phonetics* 25, 371–400.

Peterson, Gordon & Ilse Lehiste. 1960. Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America* 32(6), 693–703.

Pierrehumbert, Janet & David Talkin. 1992. Lenition of /h/ and glottal stop. In Gerard J. Docherty & D. Robert Ladd (eds.), *Papers in Laboratory Phonology II*, 90–117. Cambridge: Cambridge University Press.

Pike, Kenneth L. 1948. *Tone languages*. Ann Arbor, MI: University of Michigan Press.

Ratliff, Martha. 1987. Tone sandhi compounding in White Hmong. *Linguistics of the Tibeto-Burman Area* 10, 71–105.

Ratliff, Martha. 1992. *Meaningful tone: A study of tonal morphology in compounds, form classes, and expressive phrases in White Hmong*. DeKalb, IL: Northern Illinois University Press.

Ratliff, Martha. 2010. *Hmong-Mien language history*. Canberra: Pacific Linguistics.

Riddle, Elizabeth M. 1994. Relativization, parataxis, and underspecification in White Hmong. In Thomas J. Hudak & Karen L. Adams (eds.), *Papers from the Second Annual Meeting of the Southeast Asian Linguistics Society*, 263–278. Phoenix, AZ: Arizona State University, Program for Southeast Asian Studies.

Seyfarth, Scott & Marc Garellek. 2018. Plosive voicing acoustics and voice quality in Yerevan Armenian. *Journal of Phonetics* 71, 425–450.

Sjölander, Kåre. 2004. Snack sound toolkit. KTH Stockholm, Sweden. https://www.speech.kth.se/snack/.

Sleeper, Morgan. 2020. Contact effects on voice-onset time (VOT) in Patagonian Welsh. *Journal of the International Phonetic Association* 50, 153–167.

Shue, Yen-Liang, Patricia Keating, Chad Vicenik & Kristine Yu. 2011. VoiceSauce: A program for voice analysis. *Proceedings of the 17th International Congress of Phonetic Sciences* (ICPhS XVII), Hong Kong, 1846–1849.

Smalley, William A. 1976. The problem of consonants and tone: Hmong (Meo, Miao). In William A. Smalley (ed.), *Phonemes and orthography: Languages planning in ten minority languages in Thailand* (Pacific Linguistics Series no. 43), 85–123. Canberra: Australian National University.

Smalley, William A., Chia Koua Vang & Gnia Yee Yang. 1990. *Mother of writing: The origin and development of a Hmong Messianic Script*. Chicago, IL: The University of Chicago Press.

Sonderegger, Morgan. 2012. *Phonetic and phonological dynamics on reality television*. Ph.D. dissertation, The University of Chicago.

Styler, Will. 2017. On the acoustical features of vowel nasality in English and French. *The Journal of the Acoustical Society of America* 142(4), 2469–2482.

Umeda, Noriko. 1975. Vowel duration in American English. *The Journal of the Acoustical Society of America* 58(2), 434–445.

Wai, Yi, Marjorie Chan & Mary Beckman. 2005. An autosegmental-metrical analysis and prosodic annotation conventions for Cantonese. In Sun-Ah Jun (ed.), *Prosodic typology I: The phonology of intonation and phrasing*. Oxford: Oxford University Press.

Wang, C. Y. & C. J. Tang. 2012. The falsetto tones of the dialects in Hubei Province. *Proceedings of the 6th International Conference on Speech Prosody* (SP-12), Shanghai, 95–98.

Xu, Yi. 1994. Production and perception of coarticulated tones. *The Journal of the Acoustical Society of America* 95, 2240–2253.

Xu, Yi. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25, 61–83.

Yao, Yao. 2009. Understanding VOT variation in spontaneous speech. *UC Berkeley Phonology Lab Annual Report* 5, 29–43.

Yu, Kristine M. & Hiu Wai Lam. 2014. The role of creaky voice in Cantonese tone perception. *The Journal of the Acoustical Society of America* 136, 1320–1333.

Yuan, Jiahong & Mark Liberman. 2014. F0 declination in English and Mandarin broadcast news speech. *Speech Communication* 65, 67–74.

Zhang, Yubin & James Kirby. 2020. The role of f0 and phonation cues in Cantonese low tone perception. *The Journal of the Acoustical Society of America* 148, EL40.

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

Zhu, Xiaonong. 2012. Multiregisters and four levels: A new tonal model. *Journal of Chinese Linguistics* 40, 1–17.