Article

Checked syllables, checked tones, and tone sandhi in Xiapu Min

Yuan Chai [1*] and Shihong Ye [2]

[1] University of California San Diego
[2] Second High School Attached to Beijing Normal University
* Correspondence: Yuan Chai

Abstract: A “checked” syllable usually refers to one ending in a short vowel and glottal constriction, which results impressionistically in “short” and “abrupt” quality. Although common in languages of the world, it is still unclear how to characterize checked syllables phonetically. In this study we investigate the acoustic features of the checked syllables in citation and sandhi tones in Xiapu Min, an underdocumented language from Xiapu, China. We conducted a production experiment and analyzed the F0, voice quality, and duration of the vowels in checked syllables. The results show that, in citation tones, checked syllables are realized with distinct F0 contours from unchecked syllables, along with a creakier voice quality in the end, and a shorter duration overall. In sandhi tones, checked syllables lose their distinct F0 contours and the syllable-final glottalization. However, the short duration of checked syllables is retained after sandhi. This study lays out the acoustic properties that tend to be associated with checked syllables and can be used when testing other language varieties. The fact that sandhi checked tones lose glottalization but preserve their shorter duration indicates that glottalization might be lost prior to duration differences when checked syllables become unchecked diachronically.

Keywords: checked syllable, checked tone, sandhi, Xiapu Min

1. Introduction

Xiapu Min is a variety of Eastern Min spoken in Xiapu County (Ningde, Fujian) China (see Figure 1 for the map). It has approximately 500000 speakers (Wen, 2015). Checked syllables in Xiapu Min are syllables that are closed by glottal stop and that bear specific tones. Xiapu Min has seven lexical tones, two of which are associated with checked syllables and will be referred to as “checked tones”: high-falling-checked (54 in Chao numerals) and low-falling-checked (21). The other five tones are associated with unchecked syllables and will be referred to as “unchecked tones”: high-level (44 in Chao numerals), low-level (11); mid and high-rising (23, 35); and falling (42) (Wen, 2015). We will henceforth refer to the high-falling-checked and low-falling checked tones using one numeral as T5 and T2 to distinguish them from unchecked tones. The goals of the paper are to clarify what it means for a phonological unit to be “checked”; to summarize the acoustic characteristics of checked and unchecked tones cross-linguistically and test those characteristics in Xiapu Min; to determine how being “checked” influences whether and how tones are neutralized in sandhi forms in Xiapu Min.

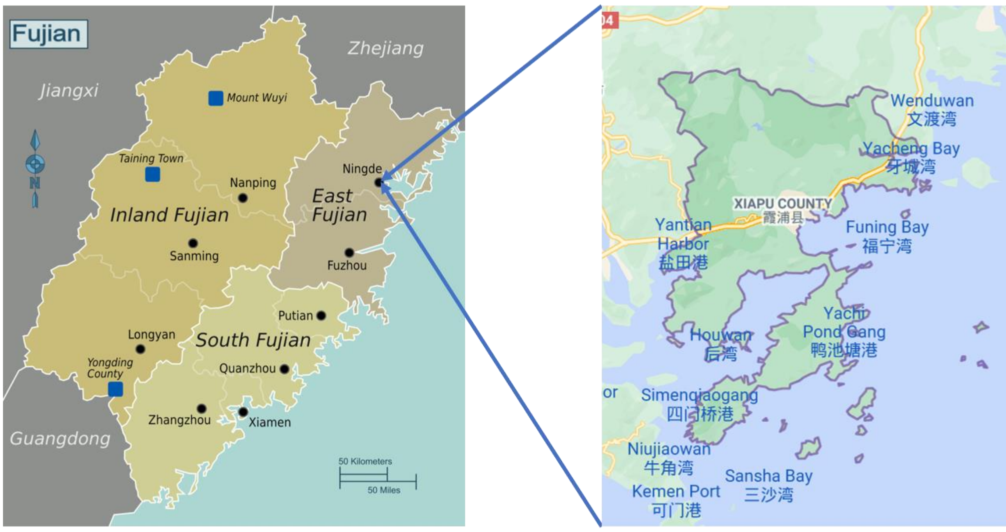

Figure 1. Map of Xiapu County. (retrieved from https://en.wikivoyage.org/wiki/File:Fujian_map.png and https://www.google.com/maps.)

A checked constituent is usually described as one that is closed with an oral or a glottal stop. Checked vowels, syllables, and tones have been widely reported in various languages of many language families (Taiwanese Min: Pan, 2005, 2017; Vietnamese: Michaud, 2004; Muak Sa-aak: Hall, 2013; Isthmus Zapotec: Pickett et al, 2010; Quiotepec Chinantec: Castellanos Cruz, 2014). However, there is a lack of consensus as to what the phonological and phonetic features of checked vowels/syllables/tones are, and whether there is a unifying phonological and/or phonetic feature of checked constituents across languages.

Researchers differ in terms of the type of phonological structure they consider to be checked. Previous literature has defined checkedness as a type of vowel phonation (i.e. *checked vowel/phonation*: Cajonos Zapotec: Nellis and Hollenbach, 1980; Betaza Zapotec: Olivares, 2009; Texmelucan Zapotec: Speck, 1978) or a type of syllable (i.e. *checked syllable*: Burmese: Gruber, 2011; Vietnamese: Kirby, 2011). Another frequently-used term is "*checked tone*", which refers to the tone(s) that are associated with checked syllables exclusively (e.g. Taiwanese Min: Pan, 2005; Burmese: Gruber, 2011). Esposito (2003, 2010) claimed that for San Ana Del Valle Zapotec, the glottal stop in CVʔ-shaped syllables is a phoneme, and "checked" was **not** a phonation type (that would contrast with modal, creaky, and breathy vowels) in that language. The reason was that the acoustic characteristics (H1–H2 and H1–F3) of the vowel in a CVʔ syllable did not differ from those of a vowel in a CV syllable. Esposito (2003, 2010) therefore used **phonetic** evidence when determining whether CVʔ is a type of phonation. In contrast, Speck (1978) claimed that, in Texmelucan Zapotec, "checked" should be a phonation type because a CVʔ-shaped syllable behaves the same as a CV-shaped syllable, but differently from a CVC-shaped syllable, in terms of third-person suffixation. The glottal stop in CVʔ syllables thus is not a phoneme, but a non-modal phonation that is realized on the vowel in an open syllable. Therefore, Speck (1978) used **phonological** evidence when determining whether CVʔ is a type of syllable or a type of phonation. In the current study, we also use **phonological** evidence to determine whether the CVʔ-shaped constituent in Xiapu Min is a syllable or a phonation type of vowels, because ultimately we are interested in how "checked" units differ phonologically from unchecked ones.

We propose that "checked" in Xiapu Min is a syllable type opposed to unchecked syllables, and define the tones that are designated to checked syllables as "checked tones". There is phonological evidence for proposing checked syllables in Xiapu Min. First, there are CV, CVN, and CVʔ-shaped syllables in Xiapu Min, and they differ phonologically. In a disyllabic compound, when the onset of the second syllable is /t/, that onset becomes [l] when it follows a CV syllable, [n] following a CVN syllable, and remains [t] when it follows a CVʔ syllable. The phonological rules and examples are presented in Table 1. This phenomenon has been reported in Wen (2015). Rules (1)-(3) demonstrate a contrast among CV, CVN, and CVʔ in phonological transformations, indicating that CVʔ is a contrastive syllable type in Xiapu Min.

Table 1. Onset change after different types of syllables[1].

	Phonological rules	Examples
(1)	/t/ → [l] / CV___	/t^he 42 **t**ain 23/ → [t^he 55 **l**ain 23] 体重 "body weight"
(2)	/t/ → [n] / CVN___	/poŋ 44 **t**oʔ 5/ → [poŋ 44 **n**oʔ 5] 饭桌 "dining table"
(3)	/t/ → [t] / CVʔ___	/t^heʔ 5 **t**o 23/ → [t^he 55 **t**o 23] 铁路 "railroad"

[1] Note that the phenomenon in Table 1 is undergoing changes and loss. For example, Wen (2015) found that these onset change rules apply to high-frequency colloquial words, but not to the words used in a formal register.

The second evidence for classifying CVʔ as a syllable type is that it is associated with tones that are distinct from CV and CVN syllables. CVʔ can only be realized in T2 or T5, whereas CV and CVN can only be realized as T11, T23, T35, T42, and T44. Since tone is realized over the entire syllable rather than over a segment, "checked" is more likely to be a feature that is associated with the entire syllable rather than the vowel.

The third evidence is from the origin of checked syllables and tones in Xiapu Min. Checked tones in Chinese languages can be traced back to the "*ru*" or "*entering tone*" in Middle Chinese (Norman, 1988). The entering tone in Middle Chinese is associated with specific syllable type – syllables that are closed by oral stop /p, t, k/. Over time, the oral codas in some varieties of Chinese were lost and replaced by a glottal stop. Xiapu Min is one such variety. A comparison between Xiapu Min and a related variety – Taiwanese Min – illustrates that the glottal codas in Xiapu Min are derived from oral codas, and are thus more likely to be a phoneme. Table 2 shows that words with a coda of /t/ or /k/ in Taiwan Min all end with a /ʔ/ in Xiapu Min. Given that CVʔ-shaped syllables in Xiapu Min behave differently from CV and CVN-shaped syllables, carry checked tones that are distinct from unchecked syllables, and are historically derived from CVT-shaped syllables, **it is the syllable that is checked in Xiapu Min**.

Table 2. Comparison between Taiwan Min and Xiapu Min codas (Taiwanese Min data is from Chien and Jongman, 2018; Xiapu Min data is from the fieldwork by the authors).

Taiwanese Min	Xiapu Min	Gloss
/tok 53/	/tuʔ 2/	毒 "to poison"
/tok 21/	/tuʔ 5/	督 "to supervise"
/sit 53/	/θiʔ 2/	实 "concrete"
/sit 21/	/θiʔ 5/	失 "to lose"

One main goal of this study is to characterize the phonetic features of checked syllables and tones in Xiapu Min. Checkedness has been described impressionistically as being "abrupt", "short", and "choked" in quality (Donohue, 2013; Michaud, 2004; Oakden, 2017; Pan, 2005; Pickett et al., 2010). In general, three phonetic features have emerged as the most common ones for distinguishing checked (vs. unchecked) constituents: shorter duration, glottalization, and a distinct pitch contour. In terms of duration, checked syllables have been found to be shorter than unchecked syllables in Xiamen Min and Nanjing Mandarin (Lai, 2016; Oakden, 2017). In terms of glottalization, Iwata et al. (1979) measured the glottal opening of speakers when producing checked and unchecked syllables in Taiwanese Min using fiberscopy. They observed that there was adduction of the false vocal folds shortly after the oral closure of /p, t, k/, and at the early portion of vowels closed by a glottal stop. The adduction of false vocal folds indicates broader laryngeal constriction in checked syllables (Esling et al. 2019). Pan (2005) found that in Taiwanese Min, at the subject and the verb positions, low-falling checked tone has more glottal constriction (represented by lower H1–A2 value) than the low-falling unchecked tone. Gruber (2011) found that in Burmese, the checked tone had smaller open quotient than the breathy and modal tone, indicating more constriction associated with checked tone. Last, in terms of pitch, Chinese languages that have checked syllables all preserve checked tones that have distinct pitch contour as opposed to unchecked tones. Here are a few representatives from Min (闽) and Jin (晋) varieties of Chinese languages that have checked tones associated with checked syllables: Yun'ao Min (checked tones 21 and 45: Lin and Lin, 2006), Xiamen Min (checked tones 32 and 4: Lai, 2006), Jincheng Jin (checked tones 2: Li, 2014), and Wutai Jin (checked tone 33: Li, 2014).

The three phonetic features – shorter duration, glottalization, and distinct pitch contour – do not necessarily co-occur in every language that has checked constituents. Quiotepec Chinantec, an Otomanguean language from San Juan Quiotepec, Oaxaca, Mexico, allows both short and long vowels to be checked. A near-minimal pair is /t**o**ˀ 3/ "honey" vs. /t**o**ːˀ 15/ "coal" (Castellanos Cruz, 2014, p. 153). Further, Vietnamese checked tones are not associated with glottalization. Michaud (2004) found that vowels in checked syllables with checked tones (D1 (rising) and D2 (falling)) have less glottal constriction than vowels in unchecked syllables with certain unchecked tones (B1: unchecked rising; B2: unchecked falling). The checked phonation in several Zapotec varieties is not restricted to specific tonal contours, including in Sierra Juarez (Tejada, 2012), Choapan (Lyman and Lyman, 1977), San Melchor Betaza (Olivares, 2009), Isthmus (Pickett et al., 2010) and Cajonos Zapotec (Nellis and Hollenback, 1980). In those Zapotec varieties, checked vowels can be realized with any of the lexical tones in the language.

Compared to the phonetics of checked tones in citation forms, the phonetics of checked tones in sandhi forms is less studied. Pan (2017) studied the voice quality of checked tones in sandhi in Taiwanese Min. In Taiwanese Min, a tone undergoes sandhi when it is in the middle of a phrase, but will preserve its citation form in phrase-final position. Low-checked tone /3/ changes into [5] after sandhi while high-checked tone /5/ changes into [3] after sandhi. Pan (2017) found that the /p, t, k, ʔ/ codas in checked tones are sometimes deleted in both sandhi and phrase-final citation form, whereas the coda deletion happens more frequently at phrase-final citation form than in sandhi form. Checked tones had less vocal fold contact than unchecked falling tones in sandhi form (indicated by smaller Contact Quotient values measured via electroglottography. Chien and Jongman (2019) compared sandhi checked tones with citation checked tones that are phonologically neutralized (i.e. [3] /5/ vs. [3] /3/; [5] /3/ vs. [5] /5/) in Taiwanese Min. They found that the F0 height and contour of sandhi checked tones and citations checked tones were indistinguishable after neutralization.

Checked tones in Xiapu Min also undergo tone sandhi in specific environments. While Pan (2017) focused on the voice quality of sandhi checked tones and Chien and Jongman (2018) focused on the F0 of sandhi checked tones, we will investigate all three features that have been found to be related to checked constituents – duration, voice quality, and F0. Another significance of studying the phonetic feature of checked tones in sandhi forms in Xiapu Min is that it can demonstrate the phonetic consequences when checked tones change into unchecked ones. Taiwanese Min checked tones remain phonologically checked after sandhi. In contrast, Xiapu Min checked tones change into phonologically unchecked tones after sandhi. Table 3 illustrates the phonological rule involved in such a kind of tone sandhi. There are two checked tones in Xiapu Min: high-falling tone 5 and low-falling tone 2. Tone sandhi happens when two tones are juxtaposed. Low-falling checked tone 2 becomes mid-level unchecked tone 44 (Rule 4), whereas high-falling checked tone 5 becomes high-level unchecked tone 55 (Rule 5) when they are followed by another tone in compounds.

Table 3. Sandhi rules in Xiapu Min. "X" = any of the seven lexical tones in Xiapu Min (The rules are based on Wen (2015) and modified based on the fieldwork data collected by the authors).

	Phonological rules
(4)	/T2, T23, T44/ → [T44] / ___ X
(5)	/T5, T35, T42/ → [T55] / ___ X

Rule (4) in Table 3 shows that, after tone sandhi, checked T2 and unchecked T23 and T44 become phonologically neutralized as T44. Checked T5 and unchecked T35 and T42 become phonologically neutralized as T55. However, it is unclear whether the neutralization is phonetically complete. Previous studies have found that tonal sandhi can either be phonetically incomplete (e.g. Mandarin T213-T35 neutralization: Kuang, 2018; Mizo rising-low tone neutralization: Lalhminghlui and Sarmah, 2018) or complete (e.g. Taiwanese Min: Chien and Jongman, 2018; Fuzhou Min T44-T242-T53 neutralization: Li, 2015). Given the possible large acoustic differences between checked and unchecked tones in citation forms in Xiapu Min, we hypothesize that the checked tones which are unchecked during sandhi will retain some of their attributes of being checked in citation form. For example, following Rule (4) in Table 3, the citation checked tone 2 should be realized as unchecked 44 before another tone. But if neutralization is incomplete, it is possible that the sandhi form retains some characteristics of being checked; e.g., it may have a shorter duration, or be creakier, than the other sandhi 44 tones derived from citation tones 23.

In sum, the research questions of this study are: 1) how do checked tones differ from unchecked tones in terms of their f0 height and contour, voice quality, and duration? 2) how do checked tones differ from unchecked tones after sandhi? Are phonologically neutralized checked and unchecked tones also phonetically neuralized completely in terms of their f0, voice quality, and duration? To answer the first question, we will measure the F0, H1*–H2*, HNR, and duration of the vowels in the checked and unchecked syllables in Xiapu Min, and perform statistical analyses to determine whether they are systematically different from each other. To answer the second question, we will perform Linear Discriminant Analysis on phonologically neutralized sandhi tones using the aforementioned acoustic measures and see whether and which acoustic parameters can effectively differentiate those neutralized tones.

2. Materials and Methods

Ten native speakers of Xiapu Min (5 women) with an average age of 53.5 participated in the production experiment conducted in Xiapu, Fujian, China. The study has been approved by the Institutional Review Board of the University of California San Diego. All the participants signed a consent form and an audio recording release consent form before participating in the experiment.

The production experiment consisted of two parts. Part 1 asked participants to produce compound words that contained citation tone minimal pairs that are neutralized after sandhi. For every pair under comparison, the target syllables had the same segments but different underlying tones. The tone of the adjacent syllable remained constant so that the effect of environment tone on the target syllable was controlled for within the pair. Rule (4) {T2, T23, T44} → T44 / ___ X was tested by 13 minimal pairs that covered the contrast between every two tones of T2, T3, and T44. Rule (5) {T5, T35, T42} → T55 / ___ X was tested by 5 minimal pairs that contrasted T5, T35, and T42 and 9 minimal pairs that covered the contrast between every two tones of T5, T35, and T42. Table 4 shows sample stimuli that cover the contrast of all phonological neutralized sandhi tones. The stimuli consisted of 27 minimal pairs, containing 59 compounds in total. Every target word was embedded in a carrier phrase of /wa42 e11 kaŋ42 **TARGET** tɕja42 la44 θø11/ ("I know how to say the word **TARGET**"), and was repeated twice, resulting in 118 compound words being produced by each participant.

Part 2 asked the participants to produce minimal pairs of citation tones. The one-syllable segments were elicited after the compound words because we did not want to prime the participants with the underlying tone of the target segments in the compound. The stimuli of Part 2 had 95 target segments in total, covering all the 30 target segments that went through sandhi in the stimuli of Part 1. Every target segment was embedded in a carrier phrase of /wa42 e11 kaŋ42 TARGET tɕja42 ka44 tɕi35/ ("I know how to say the segment TARGET"), and was produced once by each participant. The complete list of stimuli can be found as Supplementary Materials.

Table 4. Stimuli of neutralized sandhi tones in compounds for the production experiment (target segment in bold).

Sandhi rule	Contrast	Example	Gloss
{T2, T23, T44} → T44 / ___ X	T2 vs. T23	/**xuʔ 2** tsɔŋ 44/ → [**xu 44** tsɔŋ 44]	服装 "clothes"
		/**xu 23** kain44/ → [**xu 44** kain44]	护工 "caretaker"
	T23 vs. T44	/**xu 23** li42/ → [**xu 44** li42]	护理 "taking care"
		/**xu 44** tɕin42/ → [**xu 44** tɕin42]	肤浅 "superficial"
	T2 vs. T44	/**tsaʔ 2** tɕi35/ → [**tsa 44** tɕi35]	杂技 "acrobatics"
		/**tsa 44** tioŋ35/ → [**tsa 44** lioŋ35]	查账 "audit"
{T5, T35, T42} → T55 / ___ X	T5 vs. T35 vs. T42	/**iʔ 5** tiaŋ23/ → [**i 55** tiaŋ23]	一定 "certain"
		/**i 35** ŋuai23/ → [**i 55** ŋuai23]	意外 "accident"
		/**i 42** xeu23/ → [**i 55** eu23]	以后 "later"

All stimuli were presented in Chinese characters on a computer screen in a random order using PsychoPy. The participants were instructed to produce the sentences shown on the screen as they would when speaking naturally. Their productions were recorded in a quiet room in Xiapu using a Shure SM10 headset microphone, amplified by a USB-powered Focusrite Scarlett 2i2 3rd Gen preamp, and using a Dell laptop with a soundcard of Realtek ALC236.

3. Results

We segmented the vowel for each target syllable. We then calculate the following acoustic parameters: f0, H1*–H2*, and Harmonic-to-Noise Ratio between 0-500Hz (HNR). F0 correlates with the pitch of the tone. Crucially, the glottalization associated with checked syllables is predicted to have both lower H1*–H2* and lower HNR, relative to a tone with modal voice.

H1*–H2* is the difference in amplitude between the first and second harmonics (corrected for formant frequencies and bandwidths to allow for cross-vowel comparisons). Compared to modal voice, lower H1*–H2* values are correlated with more laryngeal constriction, such as the kind found during creaky voice in glottalization. In contrast, compared to modal voice, higher H1*–H2* values are correlated with glottal spreading and breathiness (Klatt and Klatt, 1990, Zhang, 2016; see overview in Garellek, 2019). HNR measures spectral noise, with lower values indicating more noise, as found for both creaky and breathy voice qualities. HNR is lower in creaky voice due to increased aperiodicity, and in breathy voice due to aspiration. We use HNR < 500 Hz because this particular noise measure is especially sensitive to aperiodicity, in addition to being sensitive to aspiration. Viewed together, H1*–H2* and HNR provide a way to distinguish modal voice from breathy and creaky voice (Garellek, 2019).

3.1. The acoustic features of checked tones in citation form

The recording of one participant's production of the citation forms was corrupted and thus discarded. One participant added an epenthetic vowel at the end of all the target words so that their recording was discarded. Eight participants produced 760 tokens in total (95 segments * 8 participants). 84 tokens were excluded because of either corrupted recording or mispronunciation, leaving 676 tokens valid for analysis. F0, H1*–H2*, and HNR were calculated using VoiceSauce (Shue et al., 2011) over every millisecond of the vowel, resulting in 148342 data points. 15 data points were excluded because their energy value was 0. VoiceSauce failed to calculate an H1*–H2* value for 2226 points and failed to calculate an HNR value for 455 points. All remaining data were z-scored by speaker. Z-scores were used in all statistical analyses to reduce between-speaker variability. Tokens with a z-score exceeding 3 were considered outliers (perhaps from tracking errors) and discarded from the data. There were 1855 outliers for F0, 1062 for H1*–H2*, and 355 for HNR. Also, the H1*–H2* values of the points whose F0 values were outliers was excluded because the former requires correct estimation of the latter. In sum, there were 146472, 143192, and 147517 data points included for f0, H1*–H2*, and HNR respectively. In order to normalize for duration differences when analyzing F0, H1*–H2*, and HNR, the data points were divided into nine equal time intervals and the mean of each interval was calculated, resulting in 6013, 5987, and 6079 normalized time points for F0, H1*–H2*, and HNR. 676 word tokens had valid acoustic measurements. The numbers of word tokens for T2, T5, T11, T23, T35, T42, and T44 are 110, 130, 82, 64, 76, 103, and 111, respectively.

3.1.1. F0

Figure 2 shows the F0 value (in st) of each tone over nine equal-timed intervals after averaging across all tokens and all speakers. The checked T5 and T2 are represented with dotted lines. The checked T5 and T2 both have a falling contour, which is similar to the contour of the unchecked falling T42. The overall F0 value of T5 is higher than T42 whereas T2 is lower than T42. Besides the unchecked falling T42, there are four other unchecked tones: a mid-level T44, a low-level T11, a mid-rising T35, and a low-rising T23. Based on the semitone values shown in Figure 2, T44 could be relabeled as T33; T35 as T24; and T23 as T13. For the purpose of consistency, the following sections will continue using the tone values proposed in the beginning of this paper without modification.

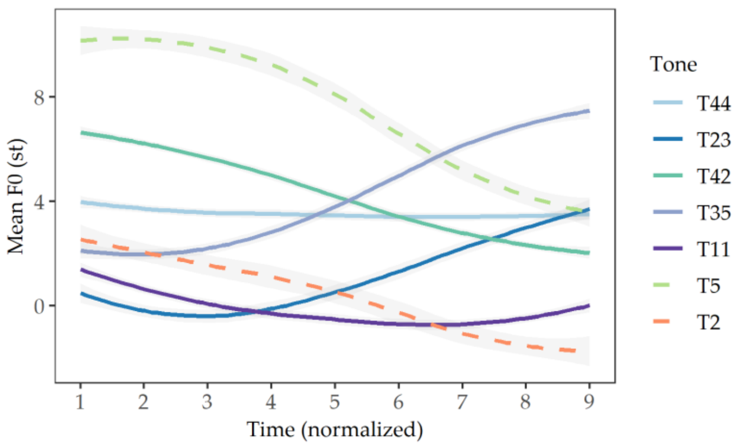

Figure 2. Average F0 track of the seven tones in Xiapu Min (in semitones with each speaker's mean F0 in Hertz of T11 as the base).

3.1.2. Voice quality

- H1*–H2*

Figure 3 shows the raw H1*–H2* values of each tone over nine equal-timed intervals after averaging across all tokens and all speakers. The checked T5 and T2 are represented with dotted lines. The graph shows that the two checked tones have a clear falling H1*–H2* contour as time proceeds, whereas the unchecked tones have a flatter H1*–H2* contour over time. The checked tones also end in a lower H1*–H2* value than the unchecked tones. We conducted a linear mixed effect analysis to test whether checked T5 and T2 have a more negative slope and end in a significantly lower H1*–H2* value than the unchecked tones. The dependent variable is H1*–H2*. The independent variables are Time, Tone, and their interaction. Participant is the random intercept. Time was centered at the last time point (Point 9) so that the intercept values reflect the H1*–H2* value of each tone at the end of the vowel. The models were implemented with the lmer() function in the lme4 package in R (Bates et al., 2015). R code for the duration model is in (1). Model (1) was run twice, once with T5 and once with T2 as the reference level of Tone. The alpha level was adjusted to 0.025 (0.05/2).

$$\mathrm{lmer}(\text{H1*–H2* (z-score)} \sim \text{Time} + \text{Tone} + \text{Time} * \text{Tone} + (1 \mid \text{Participant})) \tag{1}$$

The statistics of Model (1) are presented in Table A1 and A2 in Appendix. The results show that, for both T5 and T2, their H1*–H2* value at the end of the vowel (Point 9) is significantly lower than other vowels. Both T5 and T2 have a negative time slope on H1*–H2* (T5: -0.122; T2: -0.163), and their time slopes are significantly more negative than other unchecked tones. This indicates that T5 and T2 have a falling H1*–H2* contour whereas unchecked tones have a flatter H1*–H2* contour. Checked tones are produced with more glottal constriction at the end of the vowel than unchecked tones.

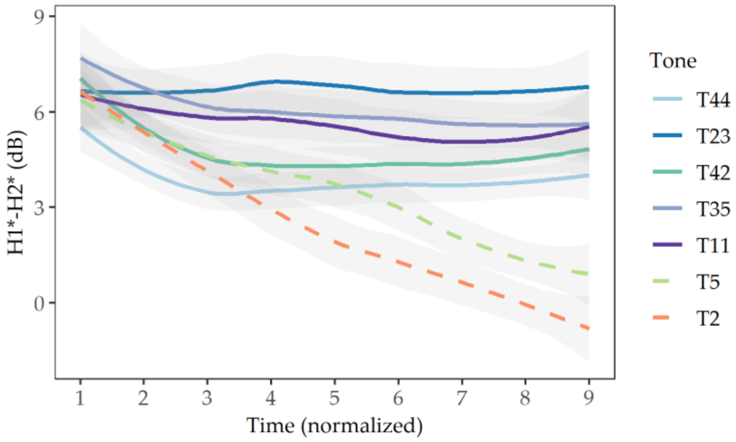

Figure 3. Average H1*–H2* track of the seven tones in Xiapu Min.

- HNR

Figure 4 shows the raw HNR values of each tone over nine equal-timed intervals after averaging across all tokens and all speakers. The checked T5 and T2 are represented with dotted lines. The HNR contour of checked T2 is below all other tones at every time point. The HNR contour of checked T5 is not distinguished from unchecked tones in its height, but there is a sudden drop in the height between Point 3 and Point 4 of the vowel. We used linear mixed effect models to test whether on average, T2 has a significant lower HNR value than unchecked tones (Model (2)), and whether T5 has a steeper falling contour at the second half of the vowel (Model (3)). The dependent variable of both models was HNR and the random intercept was Participant. In Model (2), the independent variable was Tone with T2 as the reference level. In Model (3), the independent variables were Time, Time^2 (quadratic value of Time), Tone with T5 as the reference level, and the interaction between Time, Time^2, and Tone. The Time and Time^2 variables in Model (3) were centered at Time Point 5, the middle of the vowel.

$$\mathrm{lmer}(\text{HNR (z-score)} \sim \text{Tone} + (1 \mid \text{Participant})) \tag{2}$$

$$\begin{gathered}\mathrm{lmer}(\text{HNR (z-score)} \sim \text{Time} + \text{Time\^{}2} + \text{Tone} + \text{Time} * \text{Tone} + \text{Time\^{}2} * \text{Tone} \\ + (1 \mid \text{Participant}))\end{gathered} \tag{3}$$

The statistics of Model (2) and (3) are presented in Tables A3 and A4 in Appendix. The results of Model (2) show that on average, checked T2 has a significantly lower HNR value than every other tone. The results of Model (3) show that the HNR contour of T5 is of concave shape (coefficient of Time^2 = -0.064). The derivative of T5 is positive (0.106) at Time Point 4 whereas becomes negative (-0.023) at Time Point 5. This indicates that the HNR value of T5 is decreasing in the second half of the vowel. The coefficient of Time^2 of T5 is significantly **more negative** than T23, T42, T35, and T11, and is significantly **less**

negative than T44 ($p < .001$), indicating that the HNR contour of T5 **falls faster** than T23, T42, T35, and T11, but **slower** than T44. The results of Model (2) and (3) indicates that checked T2 has a noisier quality than other tones. Checked T5 becomes increasingly noisier as it proceeds to the end of the vowel at a rate that is higher than most of the unchecked tones.

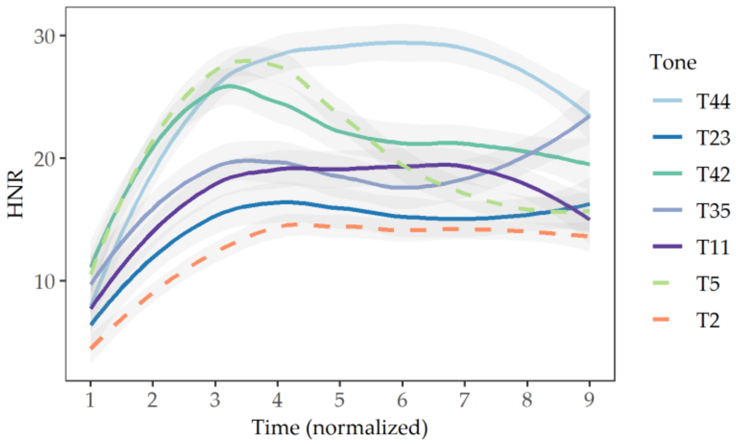

Figure 4. Average HNR track of the seven tones in Xiapu Min.

In summary, checked T2 and T5 have more glottal constriction than unchecked tones, as indicated by lower H1*-H2* values. Checked T2 has a noisier quality than unchecked tones whereas checked T5 becomes noisier abruptly in the second half of the vowel. The glottal constriction and noisy quality together indicate a creakier/more glottalized voice quality of checked tones compared to unchecked tones.

3.1.2. Duration

Figure 5 shows the duration of each tone over nine equal-timed intervals after averaging across all tokens and all speakers. To compare the duration of checked tones with unchecked tones, we used a linear mixed effect model with Duration of the vowels as the dependent variable, Tone as the independent variable, and Participant as the random intercept. The R code for the model is in (4). Model (4) was run twice, once with T5 and once with T2 as the reference level of Tone. The alpha level was adjusted to 0.025 (0.05/2). The statistics of Model (4) are presented in Tables A5 and A6 in Appendix. The results show that both checked T5 and T2 have a significantly shorter duration than every unchecked tone.

$$\text{lmer(Duration (z-score) ~ Tone + (1 | Participant))} \tag{4}$$

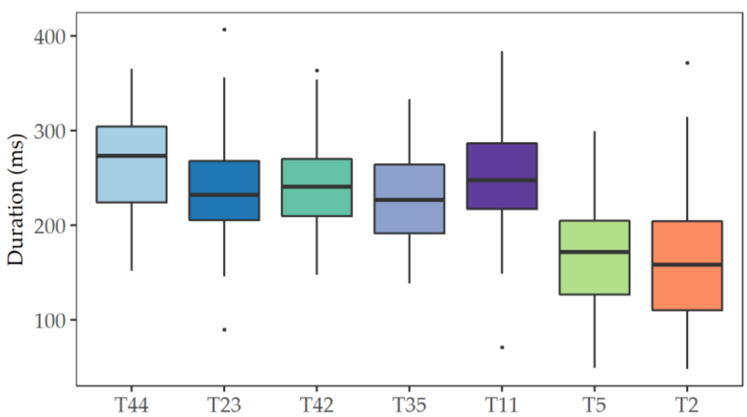

Figure 5. Average duration of the seven tones in Xiapu Min.

In summary, the two checked tones in Xiapu Min, T5 and T2, have distinct phonetic properties in all three dimensions attested. The checked tones have distinct F0 tracks compared with unchecked tones. The checked tones are produced with more glottal constrictions and aperiodicity, indicating that the vowels in checked syllables are glottalized. The glottalization gets stronger when the production proceeds towards the end of the vowel. The checked tones are shorter than unchecked tones. The phonetic features of the checked tones in Xiapu Min fit the prototypical definition of a checked constituent: **they are shorter, they end in a glottalization, and have distinct f0 values compared to unchecked tones.**

3.2. The acoustic features of checked tones in sandhi form

The segments that underwent sandhi in the compound words were the target segments for this section. Ten participants produced 1180 target segments in total (59 compounds * 2 repetitions * 10 participants). 77 segments were excluded because of either corrupted recording or mispronunciation, leaving 1103 segments valid for analyses. The acoustic measurements and the outlier exclusion process were the same as in Section 3.1. VoiceSauce yielded 123930 data points in total. 1 data point was excluded because its energy value was 0. VoiceSauce failed to calculate an H1*–H2* value for 2580 points and failed to calculate an HNR value for 940 points. There were 177 outliers for F0, 508 for H1*–H2*, and 14 for HNR. Excluding the outliers, there were 123752, 120720, and 122976 data points for F0, H1*–H2*, and HNR. The data points were divided into nine (for plotting the results) and three equal time intervals (for the linear discriminant analysis). The mean of each interval was calculated. The nine-interval data contained 9917, 9879, and 9898 normalized time points for F0, H1*–H2*, and HNR. The three-interval data contained 3307, 3305, and 3309 normalized time points for F0, H1*–H2*, and HNR. The numbers of segments for T2, T5, T23, T35, T42, and T44 are 170, 179, 153, 221, 214, and 166, respectively.

3.2.1. Neutralization among T2, T44, and T23

The Sandhi Rule (4) of Xiapu Min is {T2, T23, T44} → T44 / ___ X. It results in a neutralization between T2, T23, and T44. We conducted Linear Discriminant Analysis (LDA) (Izenman, 2013) to investigate whether the neutralized tones can be categorized by the acoustic features before and after the neutralization. LDA models use a categorical variable as the dependent variable, and use multiple parameters that can potentially differentiate the categories in the dependent variable as the independent variables. By assigning different coefficients to different parameters, the model outputs a composite linear discriminant score/scores for each token, and uses that score to classify the categories. The number of linear discriminant scores equals the number of categories in the dependent variable minus 1. For example, when there are three categories to classify, the model outputs two linear discriminant scores, which are named first and second linear discriminant scores (LD1 and LD2). The purposes of using LDA models are to compare the classification results of the model with the true categories of the data, and calculate the classification accuracy. If the classification accuracy is high, the parameters have effectively differentiated the categories in the input. The parameters that have a higher correlation with the linear discriminant scores are more effective for the classification. If the classification accuracy is at or below chance, the parameters have failed to differentiate the categories in the input. In this study, we used the proportion of the majority class as the chance level, because in random guessing, predicting all the tokens as the majority class results in the highest chance (Bosch and Paquette, 2018). The results of the LDA models can help determine whether the neutralization among the three underlyingly different tones is complete or not. The LDA models were implemented by the lda() function from the MASS package in R (Venables and Ripley, 2002).

The R code for the LDA models is in (5). The dependent variable is the citation Tone of the target segments. The independent variables are the average F0, H1*–H2*, HNR of three equal time intervals of the vowels (F0_1, F0_2, F0_3, H1*–H2*_1, H1*–H2*_2, H1*–H2*_3, HNR_1, HNR_2, HNR_3), and the Duration of the vowel. All acoustic values are the average over z-scores by each participant to reduce between-participant errors.

$$\text{lda(Tone} \sim \text{F0\_1} + \text{F0\_2} + \text{F0\_3} + \text{H1*–H2*\_1} + \text{H1*–H2*\_2} + \text{H1*–H2*\_3} + \text{HNR\_1} + \text{HNR\_2} + \text{HNR\_3} + \text{Duration)} \quad (5)$$

Figure (6a) shows the LD1 and LD2 distribution of T2, T44, and T23 in citation forms. The classification accuracy of citation forms is 94.34% and significantly higher than the 38.69% chance level ($p < .001$). We applied the LDA models on every two contrasts of T2, T44, and T23 in their sandhi forms to test the degree of neutralization between every two tones. Figure (6b) shows the LD1 distribution of each tone in each contrast. The classification accuracies of T23 vs. T44, T2 vs. T23, and T2 vs. T44 in sandhi forms are: 55.84% ($p = .147$; chance = 51.3%), 56.58% ($p = .084$; chance = 50.65%), and 77.35% ($p < .001$, chance = 51.93%). The results indicate that before sandhi, the citation forms of T2, T44, and T23 are differentiated at near-ceiling accuracy. After sandhi, T23 and T44, and T23 and T2 are completely neutralized, whereas **T2 and T44 can still be differentiated significantly above chance**.

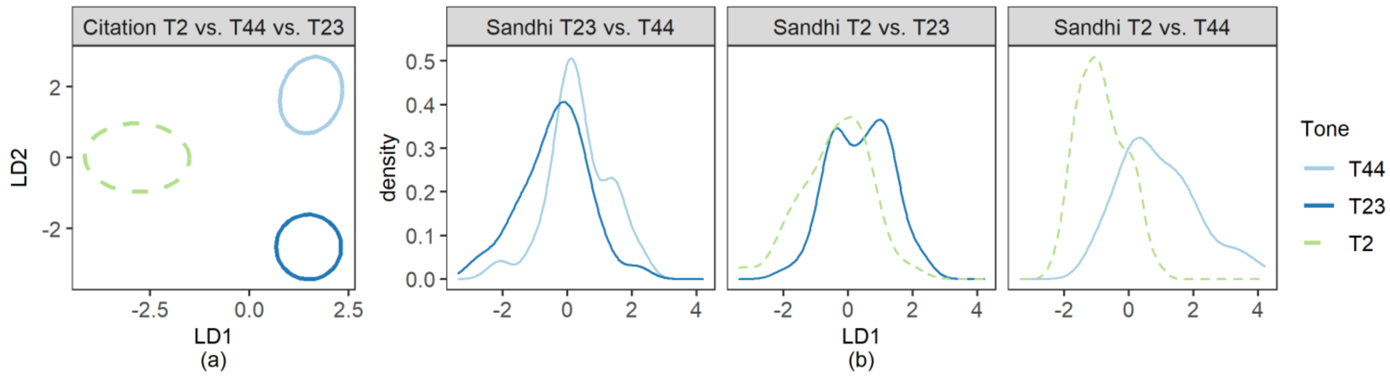

Figure 6. (a) is the first and second linear discriminant score (LD1 and LD2) distribution of T2 vs. T44 vs. T23 in citation forms. The ellipses represent 50% confidence intervals around the mean of each group. (b) is the LD1 distribution of T23 vs. T44, T2 vs. T23, and T2 vs. T44 respectively.

Next we ask what acoustic parameters contribute most to the above-chance discriminations. We correlate each acoustic parameter with the linear discriminant scores. For citation tones, LD1 explains 61% of the variance. The top three parameters that have the highest absolute correlation with LD1 are **duration, final F0, and initial HNR**. For the discrimination between T44 and T2, the top three parameters that have the highest absolute correlation with LD1 are **duration, and initial and final HNR**. The statistics of the correlation between all the parameters and the linear discriminant scores are presented in Table A7 and A8 in Appendix.

Figure 7 shows the values of F0, H1*–H2*, and HNR of T44, T23, and T2 before and after sandhi. The figure provides visual information of how the acoustic parameters change before and after sandhi. In terms of F0, the contour of three tones is well-dispersed in citation forms. After sandhi, all tones have a flat F0 contour. The F0 height of T44 is slightly lower than T23 and T2. In terms of H1*–H2*, checked T2 is produced with lower H1*–H2* than unchecked T44 and T23 in citation forms. T2 has a falling H1*–H2* contour. After sandhi, the H1*–H2* value of T2 increases and is between T44 and T23. The H1*–H2* contour of T2 is flat. In terms of HNR, the HNR of T2 is lower than T44, but similar to T23 in citation forms. After sandhi, the relation between T2, T44, and T23 remains. However, the HNR height of T2 and T23 increases. We compared the H1*–H2* and HNR of T2 before vs. after sandhi using mixed effects models, and confirmed that the increases in both parameters after sandhi are significant. The statistics are in Table A9 and A10 in Appendix. In sum, checked T2 has a constricted and noisy quality in citation form. In sandhi form, T2 becomes less constricted and less noisy, indicating the loss of glottal stop in the sandhi form. The duration of T2 is shorter than T44 and T23 both before and after sandhi. The duration of T2 is shorter in sandhi forms than in citation forms, possibly because a sandhi form is at the position of the initial syllable in a disyllabic compound word, whereas a citation form is a monosyllabic word itself.

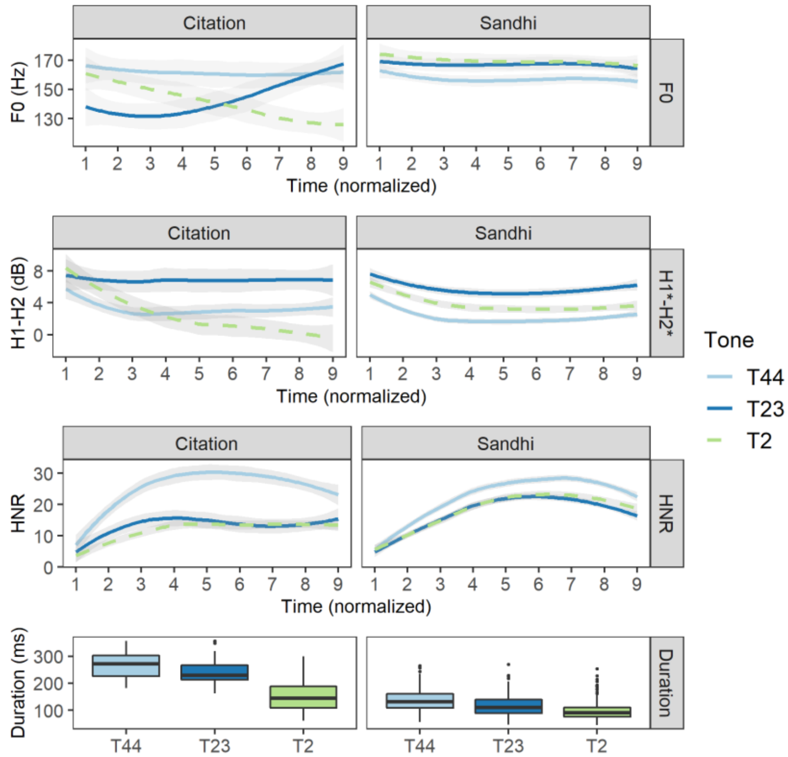

Figure 7. Acoustic parameter values of T44, T23, and T2 in citation and sandhi forms.

3.2.2. Neutralization among T5, T42, and T35

The Sandhi Rule (5) of Xiapu Min is {T5, T42, T35} → T55 / ___ X. It results in neutralization of T5, T42, and T35. Similar to Section 3.2.1, we performed LDA in this section to determine whether the neutralization between those three tones was complete or not. The R code was the same as Formula (5).

Figure (8a) shows the LD1 and LD2 distribution of T5, T42, and T35 in citation forms. The classification accuracy of citation forms is 99.06% and significantly higher than the 43.4% chance level ($p < .001$). We applied the LDA models on every two contrasts of T5, T42, and T35 in their sandhi forms. Figure (8b) shows the LD1 distribution of each tone in each contrast. The classification accuracies of T35 vs. T42, T5 vs. T35, and T5 vs. T42 in sandhi forms are: 49.85% ($p = .76$; chance = 51.61%), 75.09% ($p < .001$; chance = 50.51%), and 81.37% ($p < .001$, chance = 52.1%). The results indicate that before sandhi, the citation forms of T5, T42, and T35 are differentiated at near-ceiling accuracy. After sandhi, T35 and T42 are completely neutralized, whereas **T5 and T35, and T5 and T42 can still be differentiated significantly above chance**.

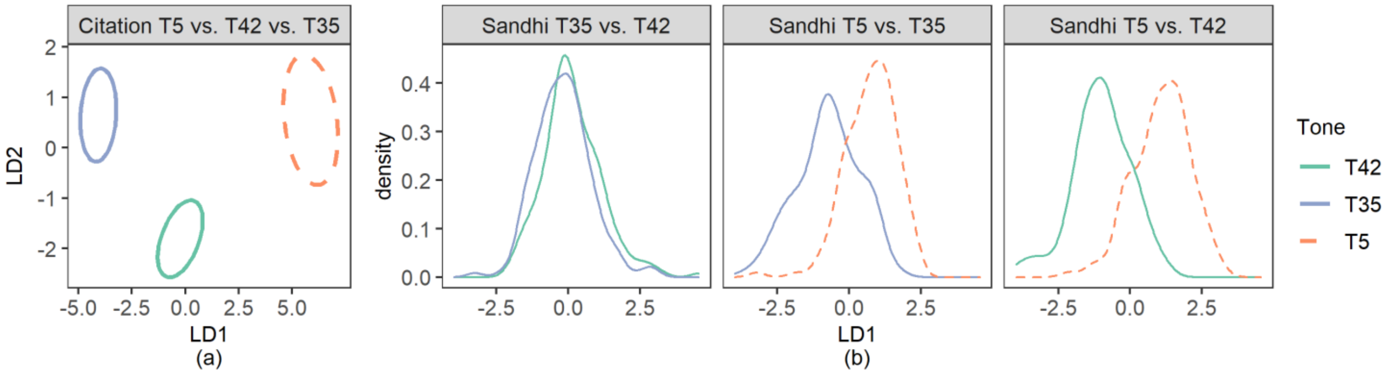

Figure 8. (a) is the LD1-LD2 distribution of T5 vs. T42 vs. T35 in citation forms. The ellipses represent 50% confidence intervals around the mean of each group. (b) is the LD1 distribution of T35 vs. T42, T5 vs. T35, and T5 vs. T42, respectively.

We correlate each acoustic parameter with the linear discriminant scores to determine which parameters contribute most to the above-chance discriminations. LD1 explains 93.86% of the variance of the citation tones. The top three parameters that have the highest absolute correlation with LD1 are **initial and mid F0, and duration**. In both discriminations between T5 and T35 and between T5 and T42 after sandhi, the top three parameters that have the highest absolute correlation with LD1 are **duration, and initial and mid HNR**. The statistics of the correlation between all the parameters and the linear discriminant scores are presented in Table A11-A13 in Appendix.

Figure 9 shows values of F0, H1*–H2*, and HNR of T42, T35, and T5 before and after sandhi. In terms of F0, the three tones have well-dispersed contours in citation forms. After sandhi, their F0 contours become flat and are largely overlapped. In terms of H1*–H2*, in citation forms, checked T5 overlaps with T42 and T35 in the first two thirds of the vowel, and has lower values than T42 and T35 in the last third. In sandhi forms, checked T5 has overall higher H1*–H2* than T42 and T35, and ends in a similar value as T42 and T35. On average, the H1*–H2* value of checked T5 has increased after sandhi. In terms of HNR, in citation forms, T5 overlaps with T42 and is higher than T35 in the first two thirds of the vowel, and has lower values than T42 and T35 in the last third. In sandhi forms, T5 has lower HNR than T42 and T35 in general. However, on average, the HNR value of T5 has increased after sandhi. In addition, before sandhi, the HNR of T5 has an abrupt fall after Point 4. After sandhi, the HNR of T5 has an overall rising contour and there is a slight fall after Point 8. The ending HNR value of T5 is higher in sandhi form than in citation form. T42 and T35 have higher HNR than T5 in sandhi forms because T42 and T35 have larger HNR increase than T5 after sandhi. We compared the H1*–H2* and HNR of T5 before vs. after sandhi using mixed effects models, and confirmed that the increases in both parameters after sandhi are significant. The statistics are in Table A14 and A15 in Appendix. In summary, checked T5 has a constricted quality and a noisy ending in citation forms. In sandhi forms, T5 becomes less constricted and less noisy, indicating the loss of glottal stop in the sandhi form. The duration of T5 is shorter than T42 and T35 both before and after sandhi. The duration of T5 is shorter in sandhi forms than in citation forms.

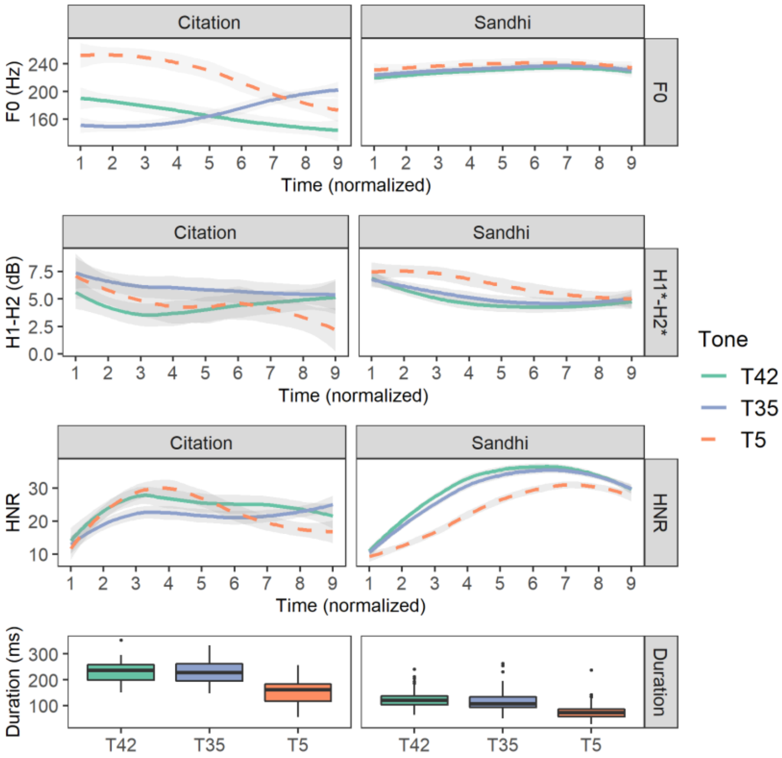

Figure 9. Acoustic parameter values of T42, T35, and T5 in citation and sandhi forms.

Table 5 presents a summary of the classification accuracy of each neutralized contrast and the top three acoustic parameters that have the highest correlation with the linear discriminant scores. Among the six neutralized pairs T23-T44, T2-T23, T2-T44, T35-T42, T5-T35, T5-T42, three of them are not completely neutralized phonetically: T2-T44, T5-T35, and T5-T42. **All those three pairs involve a checked and an unchecked tone.** The neutralizations between unchecked tones are all complete. According to the LDA results, duration is the primary cue that distinguishes checked tones from unchecked tones. HNR also appears to be an effective cue. T2 has lower HNR values than T44; and T5 has lower HNR values than T42 & T35. However, we hypothesize that this is a by-product of the short duration and the influence of onset in the checked tones. Two thirds of the target words in the stimuli have a voiceless aspirated stop (/tʰ/), voiceless affricate (/ts/), or voiceless fricative (/x, θ/) as the onset. Thus, it is possible that the aspirated and fricated onsets introduce noise into the vowels. Vowels bearing checked tones in sandhi forms are extra-short compared to those with unchecked tones (both in citation and sandhi) and to vowels with checked tones in citation forms. Sandhi checked tones are therefore likely to be more affected by the onset noise than other tokens, because their vowel duration is too short to gain periodicity after the noisy onset. Considering the artifact brought on by the onset, and the fact that average H1*–H2* and HNR values of checked tones increase after sandhi, we claim that the creaky voice quality of checked tone is largely lost in sandhi forms. The LDA results and the acoustic parameters comparisons also show that the differences in F0 between checked and unchecked tones are largely neutralized after sandhi. In summary, after sandhi, duration differentiates checked tones from unchecked tones. Their differences in voice quality and F0 are largely neutralized.

Table 5. Classification accuracies and top three parameters that have the highest correlation with LD1.

Citation/Sandhi	Contrast	Classification accuracy (chance level, *p* value)	Parameters
Citation	**T2 vs. T44 vs. T23**	**94.34%** (38.69%, <.001)	duration, final F0, initial HNR
Sandhi	T23 vs. T44	55.84% (51.3%, .147)	
	T2 vs. T23	56.58% (50.65%, .084)	
	T2 vs. T44	**77.35% (51.93%, <.001)**	duration, initial and final HNR
Citation	**T5 vs. T42 vs. T35**	**99.06%** (43.4%, <.001)	initial and mid F0, duration
Sandhi	T35 vs. T42	49.85% (50.61%, .76)	
	T5 vs. T35	**75.09% (50.51%, <.001)**	duration, initial and mid HNR
	T5 vs. T42	**81.37% (52.1%, <.001)**	duration, initial and mid HNR

4. Discussion and conclusion

In this study, we argue that Xiapu Min has checked syllables/tones, consistent with other Min languages. There is an opposition between CVʔ syllables vs. CV and CVN syllables in terms of the phonological change in the onset of the subsequent syllable. Moreover, there is an opposition between tones that are associated with CVʔ syllables vs. with CV & CVN syllables. Finally, the glottal stop in CVʔ syllables is historically derived from oral stops based on evidence from neighbouring varieties.

In this study, we also determined how checked tones in Xiapu Min differ from unchecked tones, in terms of their F0 height and contour, their voice quality, and duration. These parameters are the ones most often associated with checkedness across languages. The results show that checked tones in Xiapu Min differ from unchecked tones in all three dimensions. The two checked tones – T5 and T2 – have distinct falling contours from the unchecked tones in Xiapu Min. They are also produced with more constriction and noisier voice quality at the end of the vowel. Such evidence suggests that the ends of vowels in checked syllables in Xiapu Min are glottalized. Checked tones also have a shorter duration than unchecked tones. Thus, all three primary phonetic features of checked constituents that are found in other languages apply for Xiapu Min checked tones. We recommend that future studies on checked constituents in other languages focus on these three prototypical phonetic properties as well.

We further showed how checked tones change when they are phonologically neutralized with unchecked tones. This study finds that incomplete neutralization only happens between unchecked and checked tones. When neutralization occurs between two unchecked tones, it is complete, at least according to the measures investigated here. A possible explanation for the different degrees of neutralization is that the speakers' production of the sandhi forms is influenced by their knowledge of the citation forms. After sandhi, the acoustic parameter that most effectively differentiates checked tones from unchecked tones is duration; the F0 and voice quality differences between them in citation forms are largely neutralized. In future work, we plan to conduct perception studies that manipulate F0, voice quality, and duration separately in sound signals, and test whether the listeners are more sensitive to duration than to the other two cues when identifying whether a tone is checked or not.

The phonetic features of checked tones in sandhi forms also help clarify the relation between shorter duration and glottalization in checked tones as they occur in citation forms: are these independent correlates of checkedness, or interdependent correlates of a common articulatory goal? While short duration and glottalization might be independent targeted features for realizing a checked syllable, it is also possible that glottalization is the means of realizing a short duration of checked syllable, because glottalization can result in an abrupt "shutting off" of voicing, effectively shortening the duration of the rhyme. Another possibility is that glottalization itself is the production target, whereas the short duration is a by-product of glottalization. The results of this study provide some clarity on the matter: in sandhi checked tones, glottalization is weakened while the short duration is preserved. This suggests that glottalization might serve as a means of achieving a shorter vowel duration, which is the intended articulatory target. Once the checked syllable is followed by another syllable in compound words, glottalization is weakened while the short duration is realized by an earlier onset to the next syllable. In isolation however, there is no following sound to help keep the checked syllable short; glottalization could therefore be used in citation forms to inhibit voicing.

The hypothesis that glottalization is a medium of realizing the short duration of checked syllables can also help explain the diachronic change of the phonetic features of checked syllables in Chinese languages. There is a tendency in Chinese languages for checked syllables to become unchecked diachronically (referred as "rusheng shuhua" 入声舒化 in Chinese literature). Several studies have proposed that this process involves three steps in general: oral coda lenition to glottal coda → glottal coda deletion along with duration lengthening → loss of distinct tone contour (Cao, 2002, Gu, 2015; Shen, 2007; Song, 2009; Xing and Meng, 2006; Yang, 1982). Zuo (2015) proposed that there is an intermediate stage between Stage 2 and 3 where the glottal stop coda is weakened to weak glottalization on vowel and the duration of checked syllables is lengthened, but still shorter than unchecked syllables. We partly agree with Zuo (2015) and propose that the second stage, glottal coda deletion and short duration lengthening, can be further divided into two stages. The glottal coda deletion happens first, and the short duration lengthening happens subsequently. The evidence for this is based on contemporary Chinese varieties that show the characteristics of each stage, and the phenomenon in Xiapu Min that sandhi checked tones lose glottal coda but preserve their short duration. As mentioned in the introduction, checked syllables in Chinese languages end in /p, t, k/ coda in Middle Chinese. There are varieties, such as Taiwanese Min, where the oral codas are preserved in checked syllables. But other varieties, such as Xiapu Min, lost the oral constriction to coda stops, leaving /ʔ/. In the final stage of oral stop lenition, /ʔ/ may be lost in checked syllables. Indeed, several varieties have completely lost the coda but have preserved the short duration and the distinct F0 contour(s) (i.e. the checked tone(s)). For example, in Nanjing Mandarin, the glottal coda in checked tone 5 is deleted. However, the checked tone is realized with a shorter duration than the unchecked tones (checked 5: 101.8 ms vs. the shortest unchecked tone 31: 212.7 ms; Oakden, 2017). Several other varieties have lost both the coda and the short duration, but still preserve the distinct F0 contour(s) associated with CVT syllables. For example, in Changsha Xiang, words bearing Tone 24 all had checked tones in Middle Chinese. Tone 24 is thus referred to as a checked tone in Changsha Xiang. The checked Tone 24 has a distinct F0 contour from the unchecked tones; however, syllables with checked Tone 24 do not have oral or glottal codas. In addition, the checked Tone 24 does not have a shorter duration compared with the unchecked tones in Changsha Xiang (Li and Liu, 2006; Li, 2004). Lastly, Mandarin is an example of a variety of Chinese with no oral/glottal codas and without a tone that is uniquely associated with historically checked syllables; there are therefore no reflexes of historical checkedness in Mandarin. Comparing Taiwan Min, Xiapu Min, Nanjing Mandarin, Changsha Xiang, and Mandarin, we can summarize the diachronic trajectory of checked syllable/tone development in Chinese languages as follows: **Stage 1**: oral coda, short duration, and distinct F0 being associated with checked syllables → **Stage 2**: oral coda lenition to glottal stop → **Stage 3**: glottal stop deletion (loss of checked syllable) → **Stage 4**: loss of shorter duration → **Stage 5**: loss of distinct F0 contour (loss of checked tone). The reason why Stage 3 and 4 is often combined in literature might be because once the glottal coda is lost, there is not a stable means of achieving short duration in monosyllabic words. The transition from Stage 3 to 4 is fast and could be easily missed in documentation. Figure 10 is a schematic representation of the evolution of checkedness in Chinese. Future studies can identify what stage on this continuum do the checked syllables/tones in a language fall, and then predict how checked syllables/tones will change in the next stage.

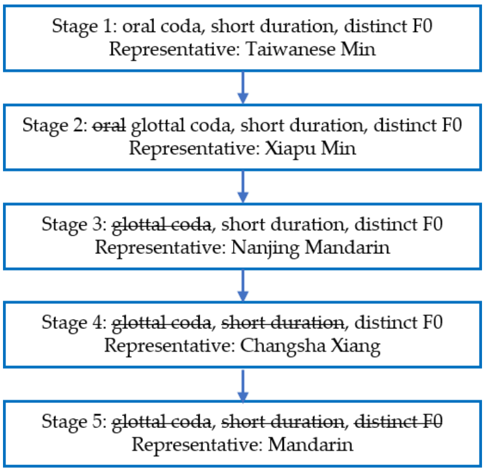

Figure 10. The proposed diachronic stages of the checked tone in Chinese languages.

Supplementary Materials: The complete stimuli wordlist of the production experiment and the sample recordings of the wordlist are at https://doi.org/10.17605/OSF.IO/M5UG2.

Author Contributions: Content hidden for double-blind review. All authors have read and agreed to the published version of the manuscript.

Funding: Content hidden for double-blind review.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of the University of XXX (protocol code: 190550; date of approval: 04/29/2019) (name hidden for double-blind review.).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the patient(s) to publish this paper.

Data Availability Statement: The sample recordings of the materials are available online at https://doi.org/10.17605/OSF.IO/M5UG2.

Acknowledgments: Content hidden for double-blind review.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A

Table A1. Results of Model (1) lmer(H1*–H2* ~ Time + Tone + Time * Tone + (1|Participant)) with T5 as the reference level

H1*–H2*	Estimate	*t*	*p*
Intercept	-0.735	-14.282	< .001
Time	-0.122	-11.195	< .001
T2	-0.287	-3.789	< .001
T44	0.530	7.125	< .001
T23	1.166	13.292	< .001
T42	0.617	8.128	< .001
T35	0.831	9.972	< .001
T11	0.853	10.522	< .001
Time * T2	-0.042	-2.613	0.009
Time * T44	0.100	6.344	< .001
Time * T23	0.121	6.547	< .001
Time * T42	0.083	5.180	< .001

Time * T35	0.071	4.059	< .001
Time * T11	0.094	5.477	< .001

Table A2. Results of Model (1) lmer(H1*–H2* ~ Time + Tone + Time * Tone + (1 | Participant)) with T2 as the reference level

H1*–H2*	**Estimate**	***t***	***p***
Intercept	-1.022	-18.130	< .001
Time	-0.163	-14.005	< .001
T5	0.287	3.789	< .001
T44	0.816	10.491	< .001
T23	1.452	16.022	< .001
T42	0.903	11.396	< .001
T35	1.118	12.930	< .001
T11	1.140	13.526	< .001
Time * T5	0.042	2.613	0.009
Time * T44	0.141	8.683	< .001
Time * T23	0.163	8.574	< .001
Time * T42	0.125	7.518	< .001
Time * T35	0.113	6.250	< .001
Time * T11	0.135	7.677	< .001

Table A3. Results of Model (2) lmer(HNR~ Tone + (1 | Participant)) with T2 as the reference level of Tone

HNR	**Estimate**	***t***	***p***
Intercept	-0.608	-19.508	< .001
Time	0.714	18.626	< .001
T5	1.145	28.781	< .001
T44	0.148	3.179	< .001
T23	0.789	19.450	.001
T42	0.500	11.341	< .001
T35	0.398	9.222	< .001
T11	-0.608	-19.508	< .001

Table A4. Results of Model (3) lmer(HNR ~ Time + Time^2 + Tone + Time * Tone + Time^2 * Tone + (1 | Participant)) with T5as the reference level

HNR	**Estimate**	***t***	***p***
Intercept	0.495	13.564	<.001
Time	0.106	8.837	<.001
Time^2	-0.064	-16.207	<.001
T2	-1.031	-20.786	<.001
T44	0.385	7.791	<.001
T23	-0.868	-14.855	<.001
T42	-0.081	-1.599	0.110
T35	-0.641	-11.599	<.001
T11	-0.490	-9.072	<.001
Time * T2	0.064	3.629	<.001
Time * T44	0.221	12.508	<.001
Time * T23	0.020	0.970	0.332

Time * T42	0.023	1.293	0.196
Time * T35	0.022	1.105	0.269
Time * T11	0.069	3.559	<.001
Time^2 * T2	0.033	5.551	<.001
Time^2 * T44	-0.023	-3.920	<.001
Time^2 * T23	0.037	5.297	<.001
Time^2 * T42	0.017	2.866	0.004
Time^2 * T35	0.053	8.074	<.001
Time^2 * T11	0.014	2.137	0.033

Table A5. Results of Model (4) lmer(Duration~ Tone + (1|Participant)) with T5 as the reference level of Tone

Duration	Estimate	*t*	*p*
Intercept	-1.285	-17.907	< .001
T2	-0.108	-1.109	0.268
T44	1.998	20.613	< .001
T23	1.439	12.558	< .001
T42	1.461	14.770	< .001
T35	1.187	10.966	< .001
T11	1.639	15.492	< .001

Table A6. Results of Model (4) lmer(Duration~ Tone + (1|Participant)) with T2 as the reference level of Tone

Duration	Estimate	*t*	*p*
Intercept	-1.393	-18.082	< .001
T5	0.108	1.109	0.268
T44	2.106	20.868	< .001
T23	1.546	13.119	< .001
T42	1.569	15.258	< .001
T35	1.295	11.579	< .001
T11	1.746	15.967	< .001

Table A7. Correlation between LD1 and LD2 in the LDA results of Model (5) for T2 vs. T23 vs. T44 in citation forms

Parameter	Correlation with LD1	Correlation with LD2
F0_1	-0.03	0.772
F0_2	0.322	0.557
F0_3	0.58	0.077
H1*–H2*_1	-0.085	-0.345
H1*–H2*_2	0.314	-0.382
H1*–H2*_3	0.436	-0.311
HNR_1	0.536	0.425
HNR_2	0.527	0.691
HNR_3	0.374	0.606
Duration	0.884	0.224

Table A8. Correlation between LD1 in the LDA results of Model (5) for T2 vs. T44 in citation forms

Parameter	Correlation with LD1

F0_1	-0.336
F0_2	-0.255
F0_3	-0.182
H1*–H2*_1	-0.172
H1*–H2*_2	-0.095
H1*–H2*_3	-0.203
HNR_1	0.625
HNR_2	0.426
HNR_3	0.521
Duration	0.928

Table A9. Results of lmer(H1*–H2* ~ Type + Tone + Tone * Type (1|Participant)) with T2 as the reference level of Tone and Citation as the reference level of Type. All H1*–H2* values are the raw values because the citation and sandhi data was transformed to z-scores separately.

H1*–H2*	Estimate	*t*	*p*
Intercept	2.578	4.459	< .001
T44	0.814	2.441	0.013
T23	4.487	12.030	< .001
Sandhi	1.441	5.394	< .001
T44:Sandhi	-2.447	-6.606	< .001
T23:Sandhi	-2.647	-6.488	< .001

Table A10. Results of lmer(HNR ~ Type + Tone + Tone * Type (1|Participant)) with T2 as the reference level of Tone and Citation as the reference level of Type. All HNR values are the raw values because the citation and sandhi data was transformed to z-scores separately.

HNR	Estimate	*t*	*p*
Intercept	10.568	5.010	< .001
T44	13.107	16.797	< .001
T23	1.581	1.811	.07
Sandhi	6.980	11.197	< .001
T44:Sandhi	-9.240	-10.638	< .001
T23:Sandhi	-2.498	-2.611	.009

Table A11. Correlation between LD1 and LD2 in the LDA results of Model (5) for T5 vs. T35 vs. T42 in citation forms

Parameter	Correlation with LD1	Correlation with LD2
F0_1	0.99	-0.02
F0_2	0.803	0.4
F0_3	-0.155	0.774
H1*–H2*_1	-0.096	0.258
H1*–H2*_2	-0.154	0.163
H1*–H2*_3	-0.247	-0.02
HNR_1	0.111	-0.325
HNR_2	0.292	-0.233
HNR_3	-0.284	-0.222
Duration	-0.722	-0.422

Table A12. Correlation between LD1 in the LDA results of Model (5) for T5 vs. T35 in citation forms

Parameter	Correlation with LD1
F0_1	0.140
F0_2	0.084
F0_3	-0.011
H1*–H2*_1	0.055
H1*–H2*_2	0.129
H1*–H2*_3	0.034
HNR_1	-0.658
HNR_2	-0.542
HNR_3	-0.274
Duration	-0.939

Table A13. Correlation between LD1 in the LDA results of Model (5) for T5 vs. T42 in citation forms

Parameter	Correlation with LD1
F0_1	0.054
F0_2	-0.042
F0_3	-0.063
H1*–H2*_1	0.137
H1*–H2*_2	0.268
H1*–H2*_3	0.176
HNR_1	-0.719
HNR_2	-0.568
HNR_3	-0.150
Duration	-0.907

Table A14. Results of lmer(H1*–H2* ~ Type + Tone + Tone * Type (1|Participant)) with T5 as the reference level of Tone and Citation as the reference level of Type. All H1*–H2* values are the raw values because the citation and sandhi data was transformed to z-scores separately.

H1*–H2*	Estimate	*t*	*p*
Intercept	4.830	8.481	< .001
T42	-0.228	-0.458	0.647
T35	1.515	3.466	< .001
Sandhi	1.527	4.192	< .001
T42:Sandhi	-1.167	-2.179	0.029
T35:Sandhi	-2.609	-5.453	< .001

Table A15. Results of lmer(HNR ~ Type + Tone + Tone * Type (1|Participant)) with T5 as the reference level of Tone and Citation as the reference level of Type. All HNR values are the raw values because the citation and sandhi data was transformed to z-scores separately.

HNR	Estimate	*t*	*p*
Intercept	20.563	8.742	< .001
T42	3.058	3.123	.002
T35	-0.247	-0.289	.773
Sandhi	2.103	2.970	.003
T42:Sandhi	3.418	3.244	.001
T35:Sandhi	5.533	5.891	< .001

References

1. Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Soft- ware* 67 (1): 1–48. https://doi.org/10.18637/jss.v067.i01.
2. Bosch, Nigel., and Luc Paquette. 2018. Metrics for discrete student models: chance levels, comparisons, and use cases. *Journal of Learning Analytics* 5 (2): 86–104. https://doi.org/10.18608/jla.2018.52.6.
3. Cao, Zhiyun. 2002. Wuhuiyu rusheng yanbiande fangshi [On the change of entering tones in Wu and Hui dialects]. *Studies of the Chinese Language* 5:441–479.
4. Castellanos Cruz, Miguel. 2014. Complejidad fonológica en el chinanteco de Quiotepec: nasalidad, foncación y tono [Phonological complexity in the Quiotepec Chinantec: nasality, phoncation and tone]. Master's thesis, Centro de Investigaciones y Estudios Superiores en Antropología Social (CIESAS), Mexico.
5. Chien, Yufan, and Allard Jongman. 2019. Tonal neutralization of Taiwanese checked and smooth syllables: An acoustic study. *Language and Speech* 62 (3): 452–474. https://doi.org/10.1177/0023830918785663.
6. Donohue, Cathryn. 2013. *Fuzhou tonal acoustics and tonology.* Lincom Europa.
7. Esling, John, Scott Moisik, Allison Benner, and Lise Crevier-Buchman. 2019. *Voice Quality the laryngeal articulator model (No. 1)*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108696555.
8. Esposito, Christina M. 2003. Santa Ana Del Valle Zapotec Phonation. Master's thesis, University of California, Los Angeles.
9. Esposito, Christina M. 2010. Variation in contrastive phonation in Santa Ana Del Valle Zapotec. *Journal of the International Phonetic Association* 40 (2): 181– 198. https://doi.org/10.1017/s0025100310000046.
10. Garellek, Marc. 2019. The phonetics of voice. In *The Routledge Handbook of Phonetics*, edited by W. F. Katz and P. F. Assmann, 75–106. New York: Routledge. https://doi.org/10.4324/9780429056253-5.
11. Gruber, James Frederick. 2011. An Articulatory, Acoustic, and Auditory Study of Burmese Tone. Doctoral dissertation, Georgetown University.
12. Gu, Qian. 2015. "Nanjing fangyan Dang Jiang liang rushengyunde gongshi bianyi [On synchronic variation and diachronic evolution of the entering tone in Dang and Jiang classes in Nanjing dialect]." *Linguistic Sciences* 14 (4): 384–393. https://doi.org/10.7509/j.linsci.201502.029722.
13. Hall, Elizabeth. 2013. A phonological description of Muak Sa-aak. *Mon-Khmer Studies* 42:26–39.
14. Iwata, R, M Sawashima, H Hirose, and S Niimi. 1979. Laryngeal adjustments of Fukienese stops: initial plosives and final applosives. *Annual Bulletin Research Institute of Logopedics and Phoniatrics* 13:61–81.
15. Izenman, Alan Julian. 2013. Linear discriminant analysis. Chap. 8 in *Modern multivariate statistical techniques*, edited by George Casella, Stephen Fienberg, and Ingram Olkin, 237–280. New York: Springer.
16. Kirby, James P. 2011. Vietnamese (Hanoi Vietnamese). *Journal of the Inter- national Phonetic Association* 41 (3): 381–392. https://doi.org/10.1017/s0025100311000181.
17. Klatt, Dennis H., and Laura C. Klatt. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America* 87 (2): 820–857. https://doi.org/10.1121/1.398894.
18. Kuang, Jianjing. 2018. The influence of tonal categories and prosodic boundaries on the creakiness in Mandarin. *The Journal of the Acoustical Society of America* 143 (6): EL509–EL515. https://doi.org/10.1121/1.5043094.
19. Lai, Wenpan. 2016. Jiyu Xiamen hua zuowei Minnan fangyan shengdiao shiyan yanjiu [Acoustic experiment on Min nan dialect using Xiamen as an example]. *Art Science and Technology* 1:30–37.
20. Lalhminghlui, Wendy, and Priyankoo Sarmah. 2018. Production and perception of rising tone sandhi in Mizo. In *Proceedings of Tonal Aspects of Languages*, 114–118. Berlin, Germany. https://doi.org/10.21437/tal.2018-23.
21. Li, Bing, and Yanni Liu. 2006. Changsha fangyan danzidiao ji biandiao de shiyan yuyinxue baogao [An acousgic of study of citation tones and tone sandhi in Changsha Chinese]. *Journal of Hunan University (Social Sciences)* 20 (4): 107–112.
22. Li, Xi. 2014. Shanxi fangyan shengdiao yanjiu [Analysis of the tone in Shanxi dialect]. Doctoral dissertation, Shaanxi Normal University.
23. Li, Xinghui. 2004. Gu rushengzi zai xiangyu zhongde fenhua [Division of the entering tone words in Xiang dialect]. *Journal of Central South Uni- versity (Social Science)* 10 (3): 394–397.
24. Li, Yang. 2015. Tone sandhi and tonal coarticulation in Fuzhou Min. In *Proceedings of the 18th International Congress of Phonetic Sciences.* Glasgow, U.K. https://doi.org/10.21437/tal.2016-25.
25. Lin, Lunlun, and Chunyu Lin. 2006. Yuedong de yige Fujian Minnan fangyandian: Yun'ao fangyan yuyin yanjiu [A southern Min dialect in the east- ern Guangdong province: a phonetic study of Yun'ao dialect]. *Fangyan [Dialect]* 1:30–37.
26. Lyman, Larry, and Rosemary Lyman. 1977. Choapan Zapotec phonology. In *Studies in Otomanguean Phonology*, edited by William R. Merrifield, 137–161. London: Summer Institute of Linguistics Publications in Linguistics.
27. Michaud, Alexis. 2004. Final consonants and glottalization: new perspectives from Hanoi Vietnamese. *Phonetica* 61:119–146. https://doi.org/10.1159/000082560.
28. Nellis, Donald G., and Barbara E. Hollenback. 1980. Fortis versus Lenis in Cajonos Zapotec Phonology. *International Journal of American Linguistics* 46 (2): 92–105. https://doi.org/10.1086/465639.
29. Norman, Jerry. 1988. *Chinese.* Cambridge, United Kingdom: Cambridge University Press.

30. Oakden, Christopher. 2017. Checked tone merger in the Nanjing dialect: an acoustic analysis. In *Proceedings of the 29th North American Conference on Chinese linguistics*, 141–152.
31. Olivares, Amador Teodocio. 2009. Betaza Zapotec phonology: segmental and suprasegmental features. Master's thesis, The University of Texas at Austin.
32. Pan, Ho-hsien. 2005. Voice quality of falling tones in Taiwan Min. In *Proceedings of INTERSPEECH 2005*, 1401–1404. Lisbon, Portugal.
33. Pan, Ho-hsien. 2017. Glottalization of Taiwan Min checked tones. *Journal of the International Phonetic Association* 47:37–63. https://doi.org/10.1017/S0025100316000281.
34. Pickett, Velma B., Maria Villalobos Villalobos, and Stephen A. Marlett. 2010. Isthmus (Juchitán) Zapotec. *Journal of the International Phonetic Association* 61:119–146. https://doi.org/10.1017/S0025100310000174.
35. Shen, Ming. 2007. "Jinyu Wutaipian rushengdiaode yanbian [On the Change of Entering Tone of Wutai Cluster in Jin Group]." *Dialect* 4:296–304.
36. Shue, Yen-Liang, Patricia Keating, Chad Vicenik, and Kristine Yu. 2011. VoiceSauce: a program for voice analysis. In *Proceedings of International Congress of Phonetic Science*, 1846–1849. Hong Kong.
37. Song, Yidan. 2009. "Nanjing fangyan zhongde rusheng housewei shiyan yanjiu [On Experiment of Entering Tone of Nanjing Dialect]." *Journal of School of Chinese Language and Culture Nanjing Normal University* 2:166– 172.
38. Speck, Charles H. 1978. The phonology of Texmelucan Zapotec verb irregularity. Master's thesis, University of North Dakota.
39. Tejada, Laura. 2012. Tone gestures and constraint interaction in Sierra Juarez Zapotec. Doctoral dissertation, University of Southern California.
40. Venables, W.N., and B.D. Ripley. 2002. *Modern Applied Statistics with S*. New York: Springer.
41. Wen, Jing. 2015. Mindongqu Xiapu fangyan yuyin yanjiu [A study on the phonology of Xiapu dialect]. Master's thesis, Fujian Normal University.
42. Xing, Xiangdong, and Wanchun Meng. 2006. Shaanbei Ganquan Yanchang fangyan rushengzi duyin yanjiu [On the pronunciation of entering tone characters in Ganquan and Yanchang dialects in northern Shaanxi]. *Chinese Language* 5:444–480.
43. Yang, Shuzu. 1982. Shanxi fangyan rushengde xianzhuang jiqi fazhan qushi [Synchronic distribution of entering tone in Shanxi dialects and their evolution tendency]. *Linguistic Research* 1:130–133.
44. 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 (4): 1493–1507. https://doi.org/10.1121/1.4944754.
45. Zuo, Fuguang. 2015. Sichuan Kejiahua yu Jiangjinhua Yibinhua rusheng yan- bian bijiao [A Comparative Study of Evolution of Entering Tone Be- tween Sichuan Hakka and Jiangjin Yibin Dialects]. *Journal of Leshan Normal University* 30 (7): 37–44. https://doi.org/10.16069/j.cnki.51-1610/g4.2015.07.009