Article

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

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

[1] University of California San Diego; yuc521@ucsd.edu
[2] Second High School Attached to Beijing Normal University; bnumiyaye@163.com
* Correspondence: Yuan Chai

Abstract: A "checked" syllable usually refers to one with a short vowel and an oral or glottal coda, which results impressionistically in a "short" and "abrupt" quality. Although common in languages of the world, it is unclear how to characterize checked syllables phonetically. In this study we investigate the acoustic features of the checked syllables in citation and sandhi forms in Xiapu Min, an underdocumented language from China. We conducted a production experiment and analyzed the F0, phonatory 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 glottalization 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 in sandhi forms. This study lays out the acoustic properties that tend to be associated with checked syllables and can be used when testing checked syllables in other language varieties. The fact that in Xiapu Min, sandhi checked tones lose glottalization but preserve their shorter duration suggests that, when checked syllables become unchecked diachronically, glottalization might be lost prior to duration lengthening.

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). There are 475,936 residents in Xiapu County in 2021 (Xiapu Government, 2021). Checked syllables in Xiapu Min are syllables that are closed by glottal stop and carry specific tones. Xiapu Min has seven lexical tones, two of which are associated with checked syllables, and will be referred to as "checked tones". They are high-falling-checked T54 (in Chao numerals, Chao, 1930) and low-falling-checked T21. The other five tones are associated with unchecked syllables, and will be referred to as "unchecked tones". They are high-level T44, low-level T11, mid and high-rising T23, 35, and falling T42 (Wen, 2015). Figure 2 shows the f0 contour of /θi/ in seven tones in Xiapu Min produced by a female native speaker. 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 summarize the acoustic characteristics of checked and unchecked tones cross-linguistically and test whether those characteristics apply to the checked tones of Xiapu Min; to determine whether the contrast between checked vs. unchecked tones are neutralized in sandhi forms; and to predict how "checked" syllables in Xiapu Min might change in the future, as a result of secondary cue loss.

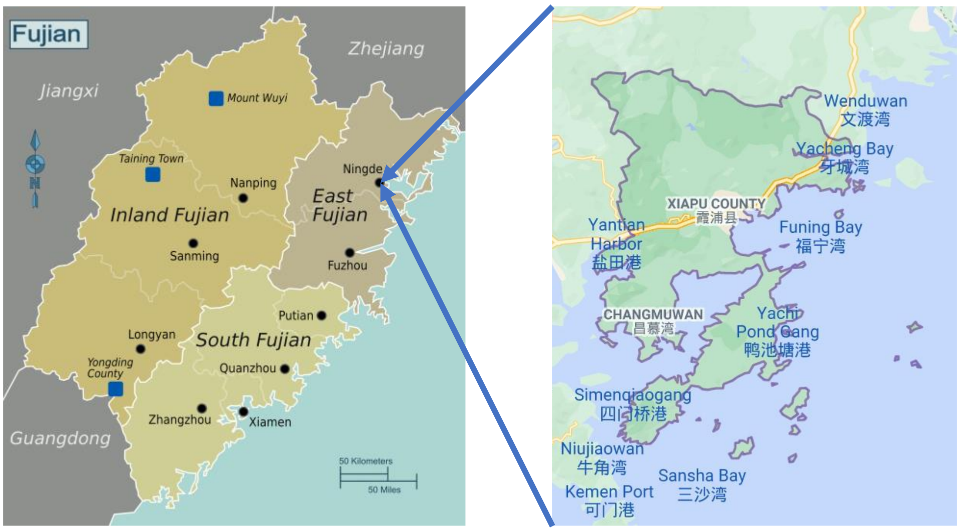

Figure 1. Map of Xiapu County. (retrieved from https://en.wikivoyage.org/wiki/File:Fujian_map.png and https://www.google.com/maps.)

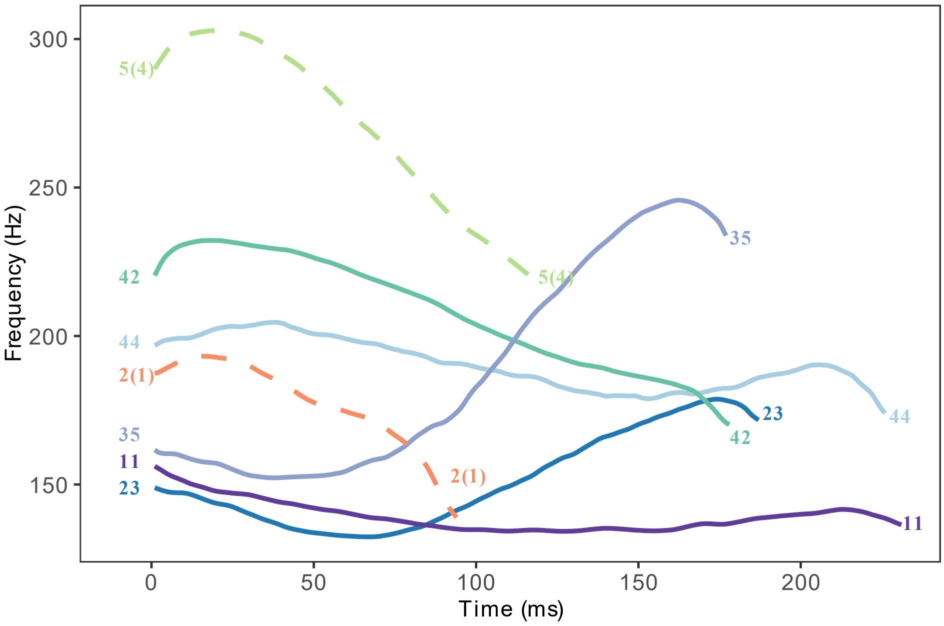

Figure 2. F0 track of /θi/ in seven tones by a female speaker.

The term "checked" has heterogeneous definitions among different languages of different families. One definition that this study will not use is "checked syllable" as the equivalence of "closed syllable" (e.g. Hall, 1971; Trask, 2004). We focus on checked syllables that are closed with an oral or a glottal stop, and also form a prosodic opposition against unchecked syllables. In Xiapu Min, two lexical tones are restricted to Vʔ syllables whereas the other five lexical tones are restricted to V and VN syllables. This is the evidence that Vʔ syllables are in prosodic opposition to V and VN syllables. Note that we consider that Xiapu Min has Vʔ checked syllables associated with checked tones, rather than having glottalized tones realized on V syllables because there is phonological evidence that glottal stop is a segment in the language. In Xiapu Min, Vʔ syllables contrast with V and VN syllables in onset changing in disyllabic compounds. In a disyllabic compound, the /t/ onset of the second syllable becomes [ɾ] when it follows a V syllable, [n] following a VN syllable, and remains [t] when it follows a Vʔ syllable. The phonological rules and examples are presented in Table 1. This phenomenon has also been reported in

Wen (2015). Rules (1)-(3) demonstrate a contrast among V, VN, and Vʔ in phonological transformations, indicating that Vʔ is a contrastive syllable type in Xiapu Min. With the evidence of tone restriction and onset changing pattern, Vʔ in Xiapu Min fits our definition of "checked syllable" that will be used in this paper. Such kind of checked syllables are widely reported in Chinese languages: Taiwanese Min (Kuo, 2013; Pan, 2016); Yun'ao Min (Zhang, 2020); Xiamen Min (Lai, 2016); Shanghai Wu (Zee & Maddieson, 1979); Nanjing Jianghuai (Sun, 2003; Yang & Chen, 2018; Chen & Wiltshire, 2013); and Meixian Hakka (Shao, 2012).

Table 1. Onset changes after different types of syllables in Xiapu Min[1].

	Phonological rules	Examples
(1)	/t/ → [ɾ] / $CV_{\sigma} + {}_{\sigma}$___	/tʰe 42 tain 23/ → [tʰe 55 ɾain 23] 体重 "body weight"
(2)	/t/ → [n] / $CVN_{\sigma} + {}_{\sigma}$___	/poŋ 44 toʔ 5/ → [poŋ 44 noʔ 5] 饭桌 "dining table"
(3)	/t/ → [t] / $CVʔ_{\sigma} + {}_{\sigma}$___	/tʰeʔ 5 to 23/ → [tʰe 55 to 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.

"Checkedness" is not associated with syllables exclusively. It can also be a type of phonation associated with vowels. The status of glottal stop determines whether checkedness is a phonation or a syllable type. For example, the glottal stop in Vʔ syllable of Texmelucan Zapotec is not a phoneme, but a phonation type. The evidence is that when adding possessive person marker to a noun (Table 2), nouns with a Vʔ syllable (4) behave the same as nouns with a V syllable (5), but differently from nouns with a VC syllable (6) (Speck, 1978). The Texmelucan Zapotec examples (Table 2) form a contrast with the Xiapu Min examples (Table 1).

Table 2. Person marker in possessive case in Texmelucan Zapotec (Speck, 1978, p. 14)

	Stem	Gloss	Possessive	Gloss
(4)	/juʔ/	"house"	/juʔm/	"her house"
(5)	/lo/	"face"	/lom/	"her face"
(6)	/sab/	"clothes	/sab mi/	"her clothes"

In the rest of the paper, we use the term "checked constituent" as an umbrella term when referring either to a checked syllable, tone, or phonation. This paper will not explore whether checked syllables and checked phonation types differ in their phonetic realization. Instead, we review the acoustic correlates of all kinds of purported checked constituents in the literature in order to capture the phonetic features of checkedness to the largest extent. Then we test how the phonetic correlates found in the literature behave in Xiapu Min checked syllables and tones, which provides the first acoustic analysis of Xiapu Min tone system. The data from Xiapu Min, an underdocumented and understudied language, will in return help clarify the phonetic nature of the checked constituents reported in a variety of languages. Four parameters have been found to distinguish checked constituents from unchecked ones: **F0, duration, voice quality,** and **vowel quality**. In Sections 1.1-4, we review how checked and unchecked constituents differ in each of those four parameters in the literature.

1.1. F0

Many Chinese languages have checked syllables associated with checked tones, which have a distinct F0 from unchecked tones. We summarize the tonal value of checked and unchecked tones of a few Chinese languages in Table 3. The tonal values are represented by Chao numerals and were supported by the F0 measurement in the references. In Examples (a)-(g), the F0 contour of the checked tones does not overlap with the unchecked ones, whereas in Examples (h)-(k), one of the two checked tones in each language overlaps with an unchecked tone in the F0 space.

Table 3. Tone values of checked and unchecked tones in selected Chinese languages

		Checked	Unchecked	Checked syllable coda	Source
(a)	Nanjing Jianghuai	5(5)[1]	31, 13, 22, 44	/ʔ/	Chen & Wiltshire, 2013; Yang & Chen, 2018
(b)	Meixian Hakka	32, 5(5)[1]	44, 21, 31, 52, 32	/p, t, k/	Shao, 2012
(c)	Changsha Xiang	23	33, 213, 41, 45	None	Shao, 2012
(d)	Chishan Xiang	35	55, 13, 31, 44, 33	None	Liu, 2013
(e)	Xiamen Min	32, 4(4)[1]	55, 35, 52, 31, 33	/p, t, k/	Lai, 2016
(f)	Anqing Jianghuai	44[2]	41[2], 35, 324, 52[2]	None	Tang, 2014
(g)	Wuhu Jianghuai	5(5)[1]	41[2], 14[2], 113[2], 54[2]	/ʔ/	Tang, 2014
(h)	East Hefei Jianghuai	**5(5)**[1,2,3]	31[2], **55**[2,3], 13[2], 52[2]	/ʔ/	Tang, 2014
(i)	Taiwanese Min	5(3), **3(1)**[1]	55, 33, 24, 51, **31**[3]	/p, t, k, ʔ/	Kuo, 2013; Pan, 2016
(j)	South Taiyuan Jin	**2(1)**[1], 42	**21**[3], 52, 35	/ʔ/	Jia, 2013
(k)	Fuzhou Min	24, **5(5)**[1]	**55**[3], 52, 32, 31, 342	/ʔ/	Shao, 2012

[1] The references used only one Chao numeral when describing the tone to indicate the shortness of the tone. We added another Chao numeral in parentheses based on the F0 values provided in the literature to describe the complete contour shape of the tone.

[2] There is between-speaker variation for the tonal value; see original source for details.

[3] Tones in bold are those for which checked and unchecked counterparts overlap in tonal value.

Apart from Chinese languages, there are other languages that were reported to have a checked tone(s). In Burmese, the checked tone is associated with checked syllables closed by glottal stop, and has a high-sharp-falling F0 contour that is distinct from the three other tones (Gruber, 2011). White Hmong is reported to have a checked tone that is closed by a glottal stop (Huffman, 1987; Ratliff, 2010). Esposito (2012), Garellek et al. (2013), and Garellek and Esposito (2021) called the checked tone as a "creaky" tone, and analyzed the glottalization at the end of the vowel as a suprasegmental creaky phonation (V̰). Nonetheless, the checked/creaky tone in White Hmong is reported to have a distinct mid/low-falling F0 contour from the other six tones in the language. A perception study has shown that the low-falling pitch contour is an essential cue for a tone to be identified as checked/creaky in White Hmong (Garellek et al., 2013).

There are also languages in which checked syllables do not carry a distinct pitch contour from unchecked syllables. Cantonese checked tones have the same tonal value as the unchecked level tones (55, 33, 22) (Chan, 1987; Qin & Mok, 2014). The checked phonation in San Melchor Betaza (Olivares, 2009) and Isthmus (Pickett et al., 2010) is not tone-dependent. Checked and unchecked vowels can be associated with the same set of tones in these languages.

1.2. Duration

Checked constituents have been found to be shorter than unchecked ones in a number of languages: Nanjing (Chen & Wiltshire, 2013; Yang & Chen, 2018; Wu, 2018), Hefei, Nantong (Wu, 2018), Anqing, and Wuhu (Tang, 2014) Jianghuai; Meixian Hakka (Shao, 2012); Fuzhou (Shao, 2012), Xiamen (Lai, 2016), and Taiwanese Min (Lin & Repp, 1989; Kuo, 2013); South Taiyuan Jin (Jia, 2013); Burmese (Gruber, 2011). White Hmong checked/creaky tone has shorter duration than other tones in general (Esposito, 2012), but the duration is dependent on vowel quality (Garellek & Esposito, 2021). In some languages, there is a length difference between their two checked tones. In Meixian Hakka, checked T5(5) is shorter than checked T32. In Fuzhou Min, checked T5(5) is shorter than checked T24 (Shao, 2012). Short duration has also been found to be a salient cue for checked tone identification. Controlling for other parameters, tokens with a shorter duration elicit more checked tone responses in Burmese (Gruber, 2011) and White Hmong (Garellek et al., 2013).

However, checkedness is not always associated with short duration. Checked tones in Changsha Xiang (Li and Liu, 2006; Li, 2004; Shao, 2012), Chishan Xiang (Liu, 2013), and Anqing Jianghuai (Tang, 2014) have longer duration than unchecked tones. Hong Kong Cantonese has both short checked and long checked tones. The long checked tone has similar duration as the unchecked tones in the language (269 vs. 284 ms) (Zhu et al., 2008). Checked vowels in San Melchor Betaza Zapotec are usually longer than modal vowels in open syllables (Olivares, 2009).

1.3. Quality of phonation

Several studies reported that checked tones were realized with non-modal phonation. Differences in phonatory quality can be represented by Open Quotient (OQ) measured from electroglottography output. A lower OQ represents a longer period of glottal closure and is an indicator of a glottalized quality (Klatt & Klatt, 1990). Acoustically, phonatory quality has been found to be indexed by spectral tilt (e.g. H1–H2, H1–A1, H1–A2, H1–A3) (Esposito, 2010a, b; DiCanio, 2009; Garellek & Keating, 2011) and F0 periodicity (Keating et al., 2015) (e.g. HNR, CPP, jitter and shimmer) (Garellek, 2019; Heiberger & Horii, 1982). A low spectral tilt and a high noise degree is an indication of glottalization, whereas a high spectral tilt and a high noise degree indicates breathy voice. Using measures of OQ and F0 jitter, Shao (2012) found that the checked T5 in Meixian Hakka and Fuzhou Min was more glottalized than unchecked tones, whereas the checked T32 in Meixian Hakka and checked T24 in Fuzhou Min were realized with a breathy voice. Checked T5 in East Hefei Jianghuai had a lower OQ value than unchecked tones, and its OQ decreased as it proceeded to the end of the vowel, indicating a glottalized quality (Tang, 2014). The Burmese checked tone had lower H1–H2, H1–A1, OQ, and oral airflow than unchecked tones, indicating that it had a glottalized quality (Gruber, 2011). The White Hmong checked/creaky tone had lower H1 value than unchecked tones at the last ninth portion of the vowel (Esposito, 2012), and lower CPP value on average (Garellek & Esposito, 2021), indicating a glottalized quality in the checked tone.

There are also checked constituents that are not associated with a glottalized quality. Checked syllables in Nanjing Jianghuai were found to be frequently produced without a glottal coda. The vowels in checked syllables had lower jitter values than those in unchecked syllables, suggesting that checked syllables were more modal than unchecked ones (Oakden, 2017). Checked tones in Taiwanese Min were not consistently glottalized (Pan, 2005). High checked tones had a higher OQ value than the unchecked high-falling tones in both citation and sandhi forms. Low checked tones had a higher OQ value than the unchecked low-falling tone in sandhi forms (Pan, 2016). Checked tones in Changsha Xiang (Shao, 2012) and Anqing and Wuhu Jianghuai (Tang, 2014) had a similar OQ to unchecked tones, indicating a modal quality.

1.4. Vowel quality

Yang and Chen (2018) reported that for Nanjing Jianghuai, compared with vowels in unchecked syllables, vowels in checked syllables had a higher F1 for /e, o, i, u, y/, higher F2 for /e, o, u/ in older generation, and higher F3 for /o, u, y/. They observed that the vowel quality difference was due to the glottal constriction gesture at the end of vowel. The glottal constriction caused jaw lowering and consequently a lower and fronter tongue position, resulting in a raising in the formant values. Wu (2008) measured the vowel space in Nanjing, Nantong, and Hefei Jianghuai, and found that the F1 of vowels in checked syllables was higher than unchecked ones. Front and back vowels were more concentrated to the middle position on the F2 scale. In Taiyuan Jin, the number of vowel contrast was reduced in checked syllables compared with unchecked syllables (Xia & Hu, 2016). While

V and VN syllables allowed six and five vowel contrasts respectively, Vʔ syllable only allowed two central vowels /ɐ, ə/.

In Table 4, we summarize the phonetic features of the checked constituents in all the languages mentioned in Section 1.1-4. Their phonetic nature is represented by four dimensions: whether the checked constituents have a distinct pitch contour, a shorter duration, a non-modal voice quality, and/or a different vowel quality from the unchecked constituents. If a feature was not described in the reference, we put "NA" in the table.

Table 4. Phonetic features of selected languages with checked constituents.

		Distinct pitch	Shorter duration	Non-modal voice quality	Different vowel quality	Sources
(a)	Nanjing Jianghuai	Y	Y	N	Y	Chen & Wiltshire, 2013; Yang & Chen, 2018; Oakden, 2017; Wu, 2008
(b)	Meixian Hakka	Y	Y	Y	NA	Shao, 2012
(c)	Changsha Xiang	Y	N	N	NA	Shao, 2012
(d)	Chishan Xiang	Y	N	N	NA	Liu, 2013
(e)	Xiamen Min	Y	Y	Y	NA	Lai, 2016
(f)	Anqing Jianghuai	Y	N	N	NA	Tang, 2014
(g)	Wuhu Jianghuai	Y	Y	N	NA	Tang, 2014
(h)	East Hefei Jianghuai	Y	Y	Y	Y	Tang, 2014; Wu, 2008
(i)	Taiwanese Min	Y	Y	N	NA	Guo, 2013; Pan, 2016
(j)	South Taiyuan Jin	Y	Y	NA	Y	Jia, 2013; Xia & Hu, 2016
(k)	Fuzhou Min	Y	Y	Y	NA	Shao, 2012
(l)	Nantong Jiang Huai	Y	Y	NA	Y	Wu, 2008; Song, 2016
(m)	Cantonese	N	Y/N[1]	NA	NA	Chan, 1987; Qin & Mok, 2014; Zhu et al, 2008
(n)	Burmese	Y	Y	Y	NA	Gruber, 2011
(o)	White Hmong	Y	Y	Y	NA	Esposito, 2010
(p)	San Melchor Betaza Zapotec	N	N	Y	NA	Olivares, 2009
(q)	Isthmus Zapotec	N	NA	Y	NA	Pickett et al., 2010

[1]Cantonese has both short and long checked tones.

1.5. Relation among the phonetic features of checked constituents

The review of the phonetic features of checked constituents across several languages shows that the phonetic nature of "checkedness" is multi-dimensional. Our next question is: what is the relation among the four features (i.e. F0, short duration, glottalization, and different vowel quality) observed in languages with checked constituents? Is each feature an independent articulatory target of the checked constituents, or is one feature the primary articulatory target, whereas the others are secondary? First, we hypothesize that vowel quality is less likely to be an independent target of the checked constituent production in Nanjing (Yang & Chen 2018), Hefei, and Nantong Jianghuai (Wu, 2008). Yang and Chen (2018) asserted that the raising in F1, F2, and F3 in checked syllables was due to the glottal constriction gesture at the end of vowel. Wu (2008) found that F1 and F2 values differed between checked and unchecked syllables, however, F1 and F2 were not significant predictors in logistic regression to discriminate checked from unchecked syllables. Vowel quality differences between checked and unchecked syllables in those languages are more likely to be a result of the glottal coda and/or shorter duration of the vowel. However, for Taiyuan Jin, where the six vowel contrasts in unchecked syllables are neutralized into two vowels in checked syllables (Xia & Hu, 2016), the vowel quality difference is more likely to be a phonological feature for checked syllables because a reduction in phonological contrasts is observed.

Second, we frequently observe that short duration and glottalization are lost together in checked syllables. Taiwanese Min has /p, t, k, ʔ/ coda in checked syllables, but the codas are frequently deleted, resulting in vowel lengthening and an increase in the voicing periodicity (Pan et al., 2016; Pan & Lyu, 2016). While short duration and glottalization might be independent articulatory targets for realizing a checked syllable, it is also possible that one is the primary feature while the other is the coarticulatory feature. The short-duration vowel gesture can be the primary articulatory target, whereas the glottalization is the means of realizing this target, 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 articulatory target, whereas the short duration is a by-product of glottalization. The relation between short duration and glottalization is closely related to the relation between short duration and the coda stop underlyingly. The diachronic change of checked syllables can help determine whether coda or short duration is the underlying feature of checked syllables. There is a tendency for checked syllables to become unchecked in Chinese languages. This phenomenon is referred as "rùshēng shūhuà" ("入声舒化") in Chinese literature. The multiple features of checked syllables do not disappear at the same time, but are lost in sequence. We propose that the feature that is lost in a later stage in the sound change is the more stable feature of checked syllable, and more likely to be the articulatory target of checked syllable. Zhu et al. (2008) listed three major paths leading to the loss of checkedness in Chinese languages. We schematize the different stages of each path proposed by Zhu et al. (2008) in Figure 3.

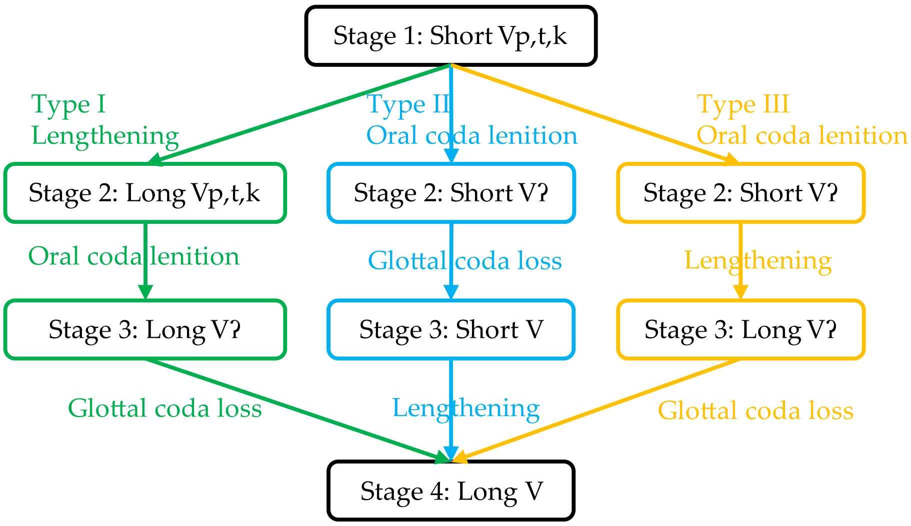

Figure 3. Three paths leading to the loss of checkedness (schematized based on Zhu et al. (2008)). (There is a mid-long duration intermediate stage in the original proposal, but is omitted here for the simplicity of illustration).

In Paths I and III, vowel lengthening happens first, and coda deletion happens next. For languages going through these two paths, the coda is more likely to be the underlying feature of checked syllables whereas the short duration is the by-product. In Path II, coda deletion happens first, and vowel lengthening happens next. For languages going through this path, the duration is more likely to be the underlying feature of checked syllables, whereas the coda is the means of reinforcing the short duration. In Nanjing Jianghuai, the glottal coda in checked T5 was frequently deleted in production. However, checked T5 was still realized with a shorter duration than the unchecked tones (T5: 101.8 ms vs. the shortest unchecked T31: 212.7 ms; Oakden, 2017). This indicates that Nanjing Jianghuai belongs to Stage 3 of Path II in the checked syllable sound change. Short duration is a more stable feature than the glottal coda for checked syllables in Nanjing Jianghuai. On the other hand, Hong Kong Cantonese has long checked syllables closed in /p/ (Zhu et al., 2008). This indicates that Hong Kong Cantonese belongs to Stage 2 of Path I in the process of checkedness loss. The coda is a more stable feature than short duration in Hong Kong Cantonese checked syllables.

Lastly, the stability of a distinct F0 contour as a feature of checked syllables can be determined by the sequencing of checked tone loss in the sound change. For example, in Changsha Xiang (Shao, 2012), Chishan Xiang (Liu, 2013), and Anqing Jianghuai (Tang, 2014), syllables that were historically closed by oral or glottal codas now become open syllables, and the vowel duration in those syllables become longer than the historically unchecked syllables. However, the distinct pitch contours that were associated with the historically checked syllables are preserved. Thus, while the coda and short duration have been lost in Changsha Xiang, Chishan Xiang, and Anqing Jianghuai, the F0 of the checked tones remains unchanged. F0 is thus a more stable feature for "checked" in those three languages than the other features. If in a language, the distinct F0 contour is lost prior to the coda deletion or the vowel lengthening, then F0 is a less stable feature for the checked syllables in that language.

Among the four acoustic features discussed above, which might be more stable for the checked syllables in Xiapu Min? Assuming that the checked syllables in Xiapu Min are also losing their checkedness as is occurring in many other Chinese languages, at what stage of the sound change might Xiapu Min currently be? What will be the next step for Xiapu Min on the path leading to the loss of checkedness? The Xiapu Min tone sandhi system makes it possible to address those questions. Xiapu Min checked tones acquire the same F0 target as unchecked tones after sandhi, indicating that checked and unchecked tones and syllables are possibly neutralized. Table 5 lists the relevant tone sandhi rules. There are two checked tones in Xiapu Min: high-falling T5 and low-falling T2. Tone sandhi happens when two tones are juxtaposed. Low-falling checked T2 becomes mid-level unchecked T44 (Rule 7), whereas high-falling checked T5 becomes high-level unchecked T55 (Rule 8) when they are followed by another tone in compounds.

Table 5. 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
(7)	/T2, T23, T44/ → [T44] / ___ X
(8)	/T5, T35, T42/ → [T55] / ___ X

Rule (7) in Table 5 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, 2016). Given the possible large acoustic differences between checked and unchecked citation tones in Xiapu Min, we hypothesize that the checked tones, which become unchecked after sandhi, will retain some of their attributes of being checked as in citation forms. For example, following Rule (7) in Table 5, the citation checked T2 should be realized as unchecked T44 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 more glottalized, than the unchecked T44 derived from other unchecked tones such as T23.

The neutralization between checked and unchecked syllables after tone sandhi can help determine the possible sound change path for Xiapu Min because it provides a synchronic condition for checked syllables to become unchecked. If one of the features of the checked syllables is retained after neutralization while the others are lost, the retained feature is likely to be more stable than others, and will be lost in a later stage in the diachronic sound change. Assuming that in Xiapu Min, checked syllables Vʔ are shorter and more glottalized than unchecked syllables, there are four alternatives for which features will be lost and which features will be retained after tone sandhi. We summarize the four alternatives in Table 6 along with the predictions that can be made based on each alternative.

Table 6. Alternatives for the possible features of checked syllables after tone sandhi

Title 1	Feature retained	Feature lost	Predicted next stage in checked syllable sound change
Alternative 1	Short duration	Glottalization	Type II Stage 3, Short V
Alternative 2	Glottalization	Short duration	Type III Stage 3, Long Vʔ
Alternative 3	Short duration & Glottalization	NA	NA
Alternative 4	NA	Short duration & Glottalization	NA

In sum, our research questions are: in Xiapu Min, 1) how do checked syllables differ from unchecked ones in terms of their F0 height and contour, phonation type, and duration? 2) how do checked syllables differ from unchecked syllables after tone sandhi? Are phonologically neutralized checked and unchecked syllables also phonetically neutralized completely in terms of their F0, phonation type, and duration? To answer the first question, we will measure the F0, H1*–H2*, HNR, duration, and vowel quality 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. The answers to Questions 1 and 2 also provide evidence to the other two inferential questions of this study. The conclusion of Question 1 answers: 3) what stage is Xiapu Min at in the checked syllable loss process? The conclusion of Question 2 potentially answers: 4) which feature/s of checked syllables is/are more stable and which is/are more likely to be lost first? In other words, what might be the next step in the checked syllable sound change for Xiapu Min?

2. Materials and Methods

2.1. Stimuli

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 XXX (name hidden for double-blind review). All the participants signed a consent form and an audio recording release consent form before participating in the experiment.

The reading material for the production experiment consists of two parts. Part 1 asked the participants to produce minimal pairs of citation tones. The stimuli of Part 1 had 95 target syllables in total. Every target syllable 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.

Part 2 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 tonal environment of the target syllable was controlled. We controlled for whether the onset of the adjacent syllable was a sonorant or obstruent, so as to minimize effects of the onset on the preceding vowel. Table 7 shows sample stimuli that display the contrasts of all phonological neutralized sandhi tones. The stimuli consisted of 21 minimal pairs, containing 41 compounds in total. Every target syllable 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. The citation forms of all the target syllables in Part 2 were covered in Part 1. The complete list of stimuli is in Supplementary Material 1. During the experiment, we elicited the compound stimuli in Part 2 first and the one-syllable stimuli in Part 1 next because we did not want to prime the participants with the underlying tone of the target syllables in the compound.

Table 7. Stimuli of neutralized sandhi tones in compounds for the production experiment (target syllable 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** kain 44/ → [**xu 44** kain44]	护工 "caregiver "
	T23 vs. T44	/**to 23** kʰeu 42/ → [**to 44** kʰeu 42]	路口 "intersection"
		/**to 44** kʰeu 42/ → [**to 44** kʰeu 42]	刀口 "wound"
	T2 vs. T44	/**tsaʔ 2** ki 44/ → [**tsa 44** ki 44]	杂技 "acrobatics"
		/**tsa 44** kaŋ 44/ → [**tsa 44** kaŋ 44]	査岗 "check up"
{T5, T35, T42} → T55 / ___ X	T5 vs. T35	/**θiʔ 5** tɕin 42/ → [**θi 55** tɕin 42]	湿疹 "eczema"
		/**θi 35** tɕia 42/ → [**θi 55** tɕia 42]	试纸 "test paper "
	T35 vs. T42	/**ka 35** kai 42/ → [**ka 55** kai 42]	价格 "price"
		/**ka 42** θe 42/ → [**ka 55** θe 42]	假设 "hypothesize"
	T5 vs. T42	/**kaʔ 5** pain 42/ → [**ka 55** pain 42]	甲板 "deck"
		/**ka 42** tsein 42/ → [**ka 55** tsein 42]	假钱 "fake money"

All stimuli were presented in Chinese characters on a computer screen in a random order using PsychoPy (Peirce et al., 2019). The participants were instructed to produce the sentences as naturally as possible. 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.

2.2. *Criteria for detecting tracking errors and outliers*

We segmented the vowel for each target syllable. All target syllables are CV or CVʔ. In syllables with a stop or fricative onset, the vowel started at the first repetitive pulse after the release of the stop or the frication noise. In syllables with a sonorant onset, the vowel started when the amplitude increased significantly. When the following word of the target syllable had a stop onset, the target vowel ended when the voicing stopped or when the formant amplitude dropped significantly, whichever came first. When the following word had a fricative onset, the target vowel ended when the frication noise started. When the following word had a sonorant onset, the target vowel ended when the amplitude decreased significantly or when the formant started to change, whichever came first.

We then calculated the following acoustic parameters: F0, F1, F2, H1*–H2*, and Harmonic-to-Noise Ratio (HNR) between 0 and 500 Hz using VoiceSauce (Shue et al., 2011). VoiceSauce calculated a value for each measurement every millisecond. F0 correlates with the pitch of the tone and was calculated using the STRAIGHT algorithm in VoiceSauce. F1 and F2 represent the height and frontness of vowels, and were calculated using PRAAT (Boersma & Weenink, 2021). The formant setting was to find 5 formants in 0-5000 Hz range. 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. 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 glottalized and breathy voice qualities. HNR is lower in creaky voice due to increased aperiodicity, and in breathy voice due to aspiration (Garellek, 2019). We use HNR measured between 0 and 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 means of distinguishing modal voice from breathy and glottalized voice (Garellek, 2019). We predict that in Xiapu Min checked syllables, the glottal coda triggers glottalization at the end of the vowel. Consequently, we predict that this creaky voice will be reflected by lower H1*–H2* and lower HNR, relative to a tone with modal voice (Garellek, 2019; Seyfarth & Garellek, 2018).

The tracking errors and outliers in the output by VoiceSauce were detected by visual inspection and statistical analysis. First, tokens whose energy value was either zero or failed to be calculated by VoiceSauce were excluded for the analyses of all acoustic measures. Next, we performed visual inspections for the F0 values in the output. Pitch tracking errors are more likely to occur when there is a non-modal voice. Thus, we manually checked the F0 output from VoiceSauce for every Vʔ token. We drew an F0 track for every Vʔ token, and inspected whether there was pitch halving or doubling in the F0 track. When the pitch tracking failed, we excluded their F0 values from F0 analysis. These files are also excluded from H1*–H2* analysis because the correct estimation of H1*–H2* depends on a correct estimation of F0. The pitch track plots and the excluded tokens are in Figure S1 and Table S1 in Supplementary Material 3.

We also performed visual and statistical inspection for the formant outputs to exclude tracking errors from formant analysis. H1*–H2* were calculated based on vowel formant. Thus, tokens with formant tracking errors were also excluded from H1*–H2* analysis. Within each vowel category, we calculated the Mahalanobis distance (De Maesschalck et al., 2000) on the F1-F2 panel between every individual token to the mean of the category. The larger the Mahalanobis distance, the more deviant the vowel is from the center of the category, and the more likely there is a tracking error for that vowel. We followed the criterion in Garellek and Esposito (2021) and Seyfarth and Garellek (2018), and regarded tokens with a Mahalanobis distance larger than 6 as an outlier and excluded them from the analysis of vowel quality and H1*–H2*. We also plotted the mean F1 and F2 of the mid-third portion of each vowel and excluded outliers detected by visual inspection. In addition, we manually checked the spectrogram of /u/s with an F2 greater than 1500 Hz. If the F2 tracking was wrong, the /u/ token was discarded from formant and H1*–H2* analysis. The vowel formant plots and the excluded tokens are in Figures S2-3 and Table S2 in Supplementary Material 3.

After excluding the tracking errors of F0, F1, and F2, we transformed the values of F0, H1*–H2*, and HNR into z-score to reduce between-speaker variation and increase the power of the statistical analyses. We calculated the Log Z-score of F0 by first log-transforming the F0 in Hertz, then z-scoring it by speaker. We conducted a log-transformation on F0 values first because the distribution of F0 was right-skewed. The log-transformation resulted in a normal distribution of the F0 and increased the validity of the statistical analyses (Keene, 1995). Tokens with a z-score exceeding 3 were considered outliers (perhaps from tracking errors) and discarded from the analysis of that measure. Tokens with F0 outliers were excluded from H1*–H2* analysis. We used the log z-score of F0 and the z-score of HNR and duration in the statistical analyses in the following sections. We used the raw value rather than the z-score for H1*–H2*, F1, and F2. H1*–H2* went through more steps of outlier exclusion than other parameters such that the data is no longer balanced by participant or tone. There were more tracking errors for the vowel /u/ than other vowels. After excluding tracking errors for formants, F1 and F2 became unbalanced by participant and vowel. Using z-score by participant for H1*–H2*, F1, and F2 could therefore distort the data and obscure effects. The R code for detecting tracking errors and outliers and all the statistical analyses in the following sections is in Supplementary Material 4.

3. Results

3.1. The acoustic features of checked tones in citation forms

Note that in Xiapu Min, checked syllables and checked tones are always associated with each other. Therefore the results measured from the vowels in checked syllables is equivalent to the results for checked tones. For simplicity, we will refer to the results for checked syllables and checked tones together as the results for checked tones. We chose "checked tone" as the reference because there are two checked tones and need to be differentiated.

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 syllables 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 675 tokens valid for analysis. VoiceSauce (Shue et al., 2011) yielded 148198 data points in total. The tracking error and outlier detection and exclusion procedures were the same as described in Section 2.2. After data exclusion, there were 144235, 135300, 135300, 129225, and 147378 data points for F0, F1, F2, 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. The descriptive statistics of the dataset can be found in Tables S3-5 in Supplementary Material 3.

3.1.1. F0

Figure 4 shows the average F0 value of each tone over nine equal-timed intervals for female and male participants respectively. Checked T2 and 5 are represented by dotted lines. For both female and males speakers, checked T5 has the highest F0 among all tones. Checked T2 has a similar onset as the rising T35. In general, tones produced by female speakers have higher F0 values than male speakers. T44, 23, 2, 11 are in a more compressed F0 range for male speakers than female speakers.

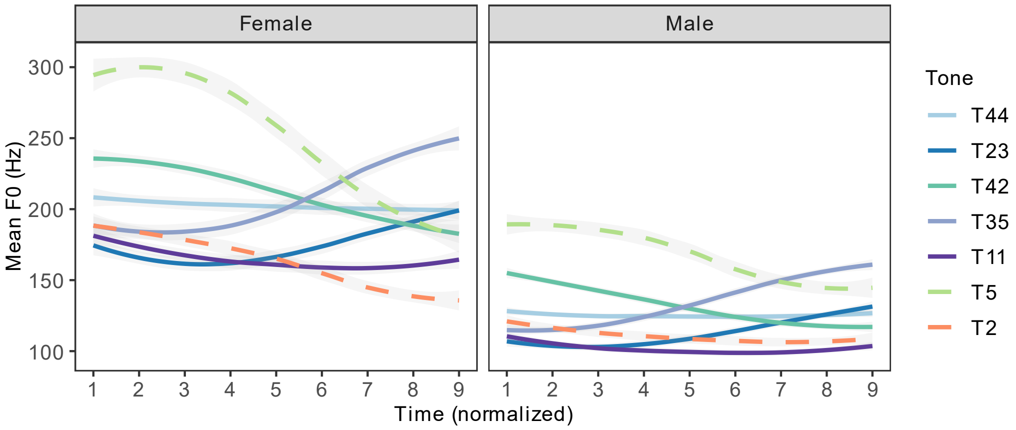

Figure 4. Average F0 track for female and male speakers of the seven tones in Xiapu Min.

Figure 5 shows that the F0 values in Hertz have large variation between females and males. We need to transform the F0 values in Hertz to a less varied scale in order to reduce between-speaker variation and establish a more uniformed representation of the tonal values. Log z-score has been found to be the most effective measure for reducing between-speaker variation among other F0 normalization methods (Zhu, 2004), and has been used in several studies (Duan & Jia, 2015; Hu et al., 2012; Jia & Li, 2012). The calculation of log z-score was described in Section 2.2. We thus use the log z-score of F0 to represent the relative pitch height and contour of the seven tones in Xiapu Min in Figure 5a.

Tonal values in Chinese languages are usually represented by Chao numerals. Shi et al. (2010) proposed a T-score to transform the log-transformed F0 value of tone into a 0-5 scale. T-scored is calculated using Formula (9). F0 represents the F0 value of the current time point. F0min and F0max are the minimum and maximum values of F0 among all time points. The correspondence between T-score and Chao numeral is in Table 8. Liu (2008) proposed that each category should overlap ±0.1 with the neighboring categories to allow flexibility. T-scores at the borderline can be assigned either the lower or the higher Chao numeral.

$$T = \frac{log_{10}(F0) - log_{10}(F0_{min})}{log_{10}F0_{max} - log_{10}F0_{min}} \times 5 \tag{9}$$

Table 8. T-score and Chao numeral conversion (Liu, 2008)

T	0-1.1	0.9-2.1	1.9-3.1	2.9-4.1	3.9-5
Chao numeral	1	2	3	4	5

Several studies on Chinese languages have adopted this T-score to transform F0 values to Chao numerals (e.g. Shao, 2012; Tang, 2014; Su, 2016). This study uses the same method to calculate the Chao numerals for Xiapu Min tones. We modified Formula (9) into Formula (10) for calculating the T-score. Rather than using the logged value of F0, we use the log z-score (LZ) of F0 because it can further reduce between-speaker variation. This modified T-score does not change the relative position of tones represented by log z-score in Figure 5a, but converts the log z-score to a 0-5 scale so that it is more convenient to assign Chao numerals to the tones. The T-scores of each tone over nine equal-timed intervals are shown in Figure 5b.

$$T = \frac{LZ(F0) - LZ(F0_{min})}{LZ(F0_{max}) - LZ(F0_{min})} \times 5 \tag{10}$$

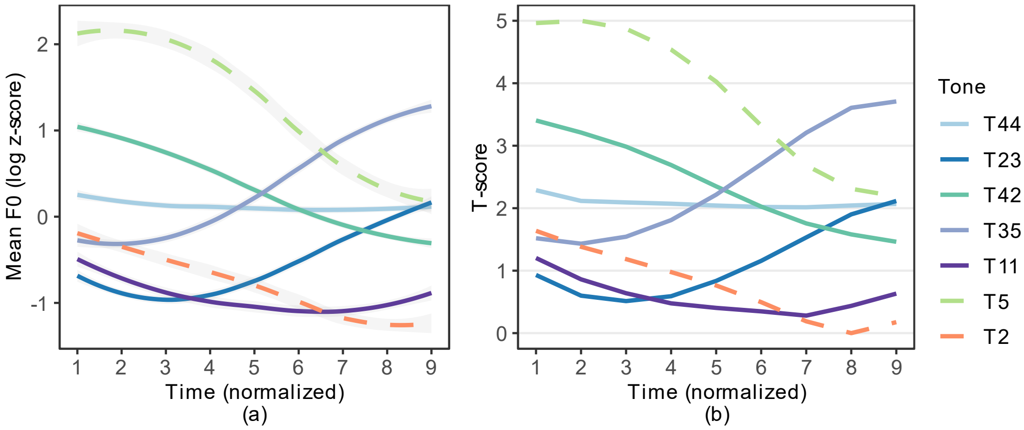

Figure 5. Log z-score (a) and T-score (b) of F0 over nine equal-timed intervals of seven tones in Xiapu Min.

The T-score at Time Point 1 and 9 of each tone and their corresponding Chao numeral based on Table 8 are listed in Table 9. Note that, based on the rules in Table 8, T11 should be referred to as 21. However, checked T2 also falls in the 21 range, and has a higher onset and a steeper fall than T11. Thus, for the purpose of differentiating the low unchecked tone from the low-falling checked tone, we assign Chao numerals 11 to T11. Our results provide an acoustic basis for the numerical value of Xiapu Min tones. We found that the reported mid-rising tone T35 does not rise as high as T5 and should be noted as 24. The reported mid-level T44 has lower onset than T42 and should be noted as 33. The Chao numeral we assigned to the tone is closer to the acoustic nature of the pitch height and shape in production. We suggest then that future studies on Xiapu Min use the Chao numerals proposed in this study. Also note that the syllables in this section were elicited in a carrier phrase. Future study should also elicit syllables in isolation and see whether the tonal value changes. For the sake of consistency, we will continue to use the original tone number throughout the rest of the study.

Table 9. T-score and Chao numerals of the seven tones in Xiapu Min (for each tone, the value for the onset is on the left and the value for the offset is on the right).

Tone (Wen, 2015)	**T5**		**T2**		**T42**		**T44**		**T35**		**T23**		**T11**	
T-score	4.96	2.19	1.64	0.18	3.40	1.46	2.29	2.07	1.52	3.71	0.93	2.12	1.20	0.63
Chao numeral	5	3	2	1	4	2	3	3	2	4	1	3	1*	1
Revised tonal values	T53		T21		T42		T33		T24		T13		T11	

* The onset of T11 should be assigned a Chao numeral of 2, but was assigned 1 to be differentiated from T2.

3.1.2. Quality of phonation

- H1*–H2*;

Figure 6 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-effects 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 model was implemented with the lmer() function in the lme4 package in R (Bates et al., 2015) (same for all other linear mixed-effects models in this paper). The R code for the H1*–H2* model is in (11). Model (11) 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).

$$\text{lmer}(\text{H1*–H2*} \sim \text{Time} + \text{Tone} + \text{Time} * \text{Tone} + (1\,|\,\text{Participant})) \tag{11}$$

The statistics of Model (11) are presented in Tables S6-7 in Supplementary Material 3. The results show that, for both T5 and T2, their H1*–H2* values at the end of the vowel (Point 9) are significantly lower than other vowels. Both T5 and T2 have a negative time slope on H1*–H2* (T5: -0.60; T2: -0.93), and their time slopes are significantly steeper 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.

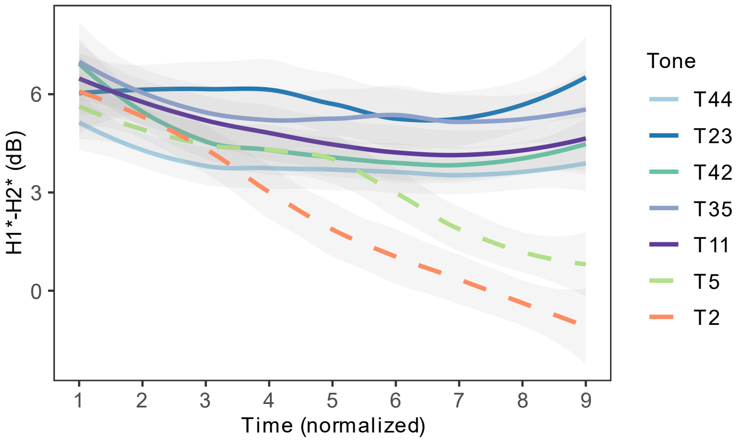

Figure 6. Average H1*–H2* track of the seven tones in Xiapu Min.

- HNR;

Figure 7 shows the raw HNR values of each tone over nine equal-timed intervals after averaging across all tokens and all speakers. The HNR contour of checked T2 is below all other tones at every time point. We used linear mixed-effects models to test whether on average, T2 and T5 have a significant lower HNR value than unchecked tones (12). Model (12) 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).

$$\text{lmer(HNR (z-score)} \sim \text{Tone} + (1 \mid \text{Participant})) \qquad (12)$$

The statistics of Model (12) are presented in Tables S8-9 in Supplementary Material 3. The results of Model (12) show that, on average, checked T2 has a significantly lower HNR value than every other tone. Checked T5 has an HNR value lower than T44, similar to 42, but higher than T2, 23, 35, and 11.

However, Figure 7 shows that there is a sudden drop in the HNR contour of T5 between Time Point 3 and 4. The drop of HNR from Time Point 4 (P4) to 9 (P9) is larger for T5 than any other tone. Table 10 shows the HNR values at P4 and P9 for all seven tones. We fitted a smooth spline for each contour using the sm.spline() function in R package "pspline" (Ramsey & Ripley, 2017), and calculated the first derivative of the contour at each time point. A positive derivative means the contour is rising. A negative derivative means the contour is dropping. A large absolute value means the rising/dropping slope is steep. The fitted spline for the HNR contour of each tone is plotted in Figure S4 in Supplementary Material 3. The HNR value predicted by the fitted spline and the first derivative at each time point for each tone are in Table S10 in Supplementary Material 3. Table 10 summarizes the HNR value at P4 and P9 and the difference between P4 and P9 for each tone. The last row is the first derivative at P4 for each tone. We see that T5 has the largest HNR fall from P4 to P9, and its first derivative at P4 has the largest negative value. Combining the evidence from visual inspection, raw HNR value difference between P4 and P9, and the negative derivative of the HNR contour, we argue that T5 has the largest HNR drop in the last two-thirds of the vowel among the seven tones. This indicates that the production of T5 targets at a noisy quality towards the end of the vowel.

Table 10. HNR value at P4 and P9, the difference in HNR between P4 and P9, and the first derivative of the HNR contour at P4.

	T5	**T2**	**T44**	**T42**	**T35**	**T23**	**T11**
P4	27.42	14.51	28.10	24.11	19.62	16.41	19.11
P9	15.54	13.47	23.19	18.96	23.31	16.28	14.64
P4-P9	**11.89**	1.04	4.91	5.15	-3.69	0.13	4.48

First derivative at P4	**-3.29**	0.67	1.46	-3.05	-0.88	0.004	0.35

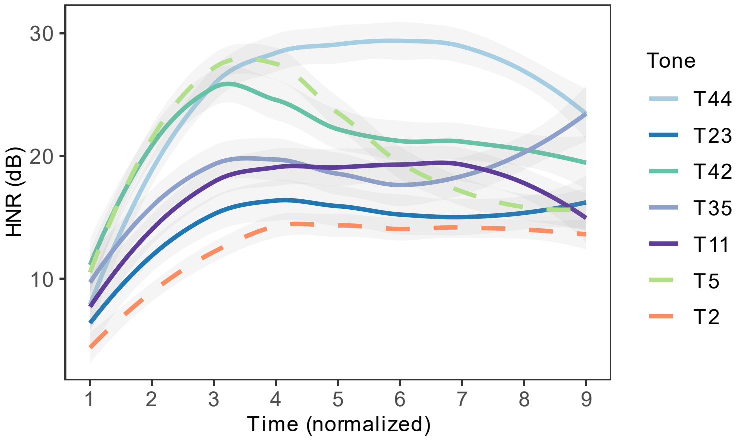

Figure 7. Average HNR track of the seven tones in Xiapu Min.

In sum, 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 last two thirds of the vowel. The glottal constriction and noisy quality together indicate a more glottalized phonatory quality of checked tones compared to unchecked tones. The HNR values do not distinguish checked from unchecked tones as consistently as the H1*–H2* values. Thus, we hypothesize that the listeners are less likely to use HNR as a salient cue in distinguishing checked tones from unchecked ones in perception. Future studies can manipulate spectral tilt and pulse periodicity separately to test the perceptual saliency of those two cues for checked tone identification in Xiapu Min.

3.1.3. Duration

Figure 8 shows the duration of each tone after averaging across all tokens and all speakers. We ran a linear regression model to compare the duration of checked tones with unchecked tones. The R code for the model is in (13). Model (13) 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 (13) are presented in Tables S11-12 in Supplementary Material 3. The results show that both checked T5 and T2 have a significantly shorter duration than every unchecked tone.

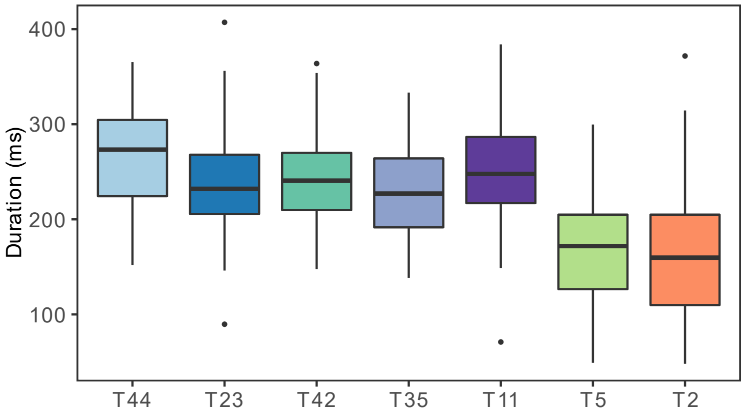

Figure 8. Average duration of the seven tones in Xiapu Min.

$$\text{lm(Duration (z-score)} \sim \text{Tone)} \tag{13}$$

3.1.4. Vowel quality

Figure 9 shows the distribution in the F1-F2 vowel formant space of the five monophthongs for checked (T5, 2) and unchecked tones (T44, 11, 23, 35, 42). We did not include diphthong for comparison to avoid the influence of formant transition in diphthong. For each token, we calculated the mean F1 and F2 of mid third of the vowel to ensure that the vowel formant is at a stable stage. Figure 9 shows that the checked and unchecked vowel ellipses have large overlaps. To determine whether checked and unchecked vowels differ in F1 and F2, we conducted linear mixed-effect analyses using models in (14) and (15). The statistics of Model (6) are presented in Tables S13-14 in Supplementary Material 3. The result shows that, for both F1 and F2, the checked tone does not differ significantly from the unchecked tone. This indicates that checked and unchecked tones do not differ in vowel quality in Xiapu Min.

$$\text{lmer(F1} \sim \text{Tone + Checkedness + (1} \mid \text{Participant))} \tag{14}$$

$$\text{lmer(F2} \sim \text{Tone + Checkedness + (1} \mid \text{Participant))} \tag{15}$$

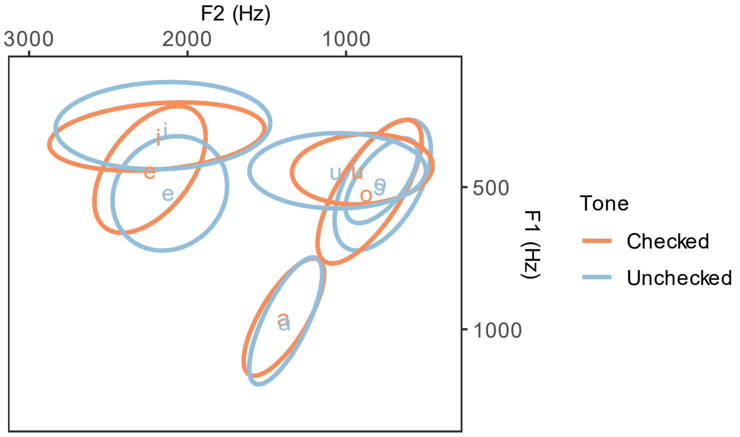

Figure 9. F1 and F2 distribution of vowels in checked and unchecked tones.

In sum, we confirm that T5 and 2 have distinct pitch values from unchecked tones, and propose a modification to the tonal values of T44, 35, and 23 based on the results from eight speakers and careful F0 normalization. We find that 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. Checked and unchecked tones are found to be different in three out of four dimensions attested: **they are shorter, they end in a glottalization, and have distinct F0 values compared to unchecked tones.** No significant differences in vowel quality have been found between checked and unchecked tones.

3.2. The acoustic features of checked tones in sandhi forms

The syllables that underwent sandhi in the compound words were the target syllables for this section. Ten participants produced 820 target syllables in total (41 compounds * 2 repetitions * 10 participants). 58 syllables were excluded because of either corrupted recording or mispronunciation, leaving 762 syllables valid for analyses. VoiceSauce (Shue et al., 2011) yielded 80360 data points in total. The tracking error and outlier detection and exclusion procedures were the same as described in Section 2.2. After data exclusion, there were 80220, 75502, 75502, 72796, and 79857 data points for F0, F1, F2, H1*–H2*, and HNR respectively. The data points were divided into nine (for plotting the results) and three equal time intervals (for the linear discriminant analysis). The descriptive statistics of the dataset can be found in Tables S3-5 in Supplementary Material 3.

3.2.1. Neutralization among T2, T44, and T23

The first sandhi rule of Xiapu Min is {T2, T23, T44} → T44 / ___ X (Table 5, Example 7). 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 & 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 & Ripley, 2002).

The R code for the LDA models is in (16). The dependent variable is the citation tone of the target syllables. 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. We did not include vowel formants in the model because no difference in vowel formants was found in the citation forms of the target syllables.

$$\begin{gathered}\text{lda(Tone} \sim \text{F0\_1} + \text{F0\_2} + \text{F0\_3} + \text{H1*–H2*\_1} + \text{H1*–H2*\_2} + \text{H1*–H2*\_3} + \\ \text{HNR\_1} + \text{HNR\_2} + \text{HNR\_3} + \text{Duration)}\end{gathered} \tag{16}$$

We compared the three tones in citation forms (T2 vs. T44 vs. T23) in the same model. Since the acoustic differences among tones in sandhi forms are likely to be largely neutralized, comparing all three tones in sandhi forms in the same model could potentially obscure the fine-grained differences. Thus, we compared every two tones in sandhi forms (T23 vs. T44, T2 vs. T23, T2 vs. T44) in three separate models. The citation tones are distinguished by two LD scores. Each pair of sandhi tones is distinguished by one LD score.

Figure (10a) shows the LD1 and LD2 distribution of T2, T44, and T23 in citation forms. The classification accuracy of the citation forms is 95% and significantly higher than the 38.75% chance level (p < .001). We applied the LDA models on each pair of the contrasts between T2, T44, and T23 in sandhi forms to test the degree of neutralization between every two tones. Figure (10b) shows the LD1 distribution of each tone in each contrast. The classification accuracies were calculated based on leave-one-out cross-validation. The classification accuracies of T23 vs. T44, T2 vs. T23, and T2 vs. T44 in sandhi forms are: 64.86% (p = .12; chance = 54.05%), 69% (p < .001; chance = 51%), and 79.31% (p < .001, chance = 51.72%). The results indicate that the citation forms of T2, T44, and T23 are differentiated at near-ceiling accuracy. In sandhi forms, however, T23 and T44 are completely neutralized, whereas **T2 and T23**, and **T2 and T44 can still be differentiated significantly above chance**. Note that T23 and T44 are tested by only one minimal pair whereas T2 vs T23 and T2 vs. T44 are tested by three and five minimal pairs respectively. The results for T23 vs. T44 may not be as representative as the other two pairs. Future studies should aim for more balanced stimuli.

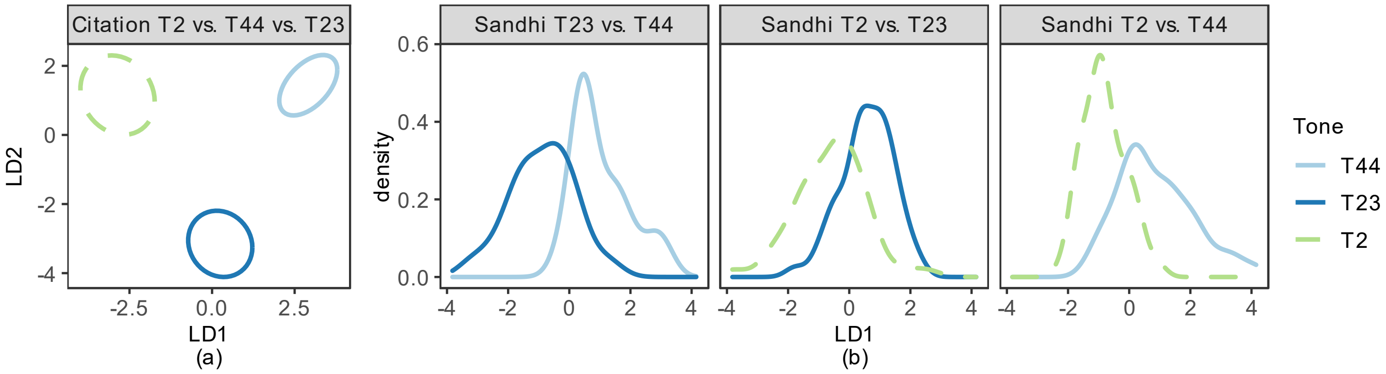

Figure 10. (a) is the first and second linear discriminant score (LD1 and LD2) distribution of T2 vs. T44 vs. T23 in citation forms. The ellipses represent 50% confidence intervals around the mean of each group. (b) is the LD1 distribution of T23 vs. T44, T2 vs. T23, and T2 vs. T44 in sandhi forms, respectively.

Next we ask which acoustic parameters contribute most to the above-chance discriminations. We correlate each acoustic parameter with the linear discriminant scores. For citation tones, LD1 explains 60.89% of the variance. The top three parameters that have the highest absolute correlation with LD1 are duration, final F0, and mid HNR. For the discrimination between T23 and T2, the top three parameters that have the highest absolute correlation with LD1 are duration, final F0, and initial H1*–H2*. 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 Model (16) and the correlations between the parameters and the linear discriminant scores are presented in Tables S15-19 in Supplementary Material 3.

Figure 11 shows the values of F0, H1*–H2*, HNR, and duration of T44, T23, and T2 in citation and sandhi forms. In terms of F0, the contours of the three tones are well dispersed in citation forms. In sandhi forms, 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. In sandhi forms, 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. In sandhi forms, the difference in HNR between those three tones remains, but becomes much smaller. The HNR of T2 and T23 increases. We compared the H1*–H2* and HNR of T2 between citation and sandhi forms using mixed-effects models, and confirmed that the increases in both parameters after sandhi are significant. The statistics are in Tables S20-21 in Supplementary Material 3. In sum, checked T2 has a constricted and noisy quality in citation forms. In sandhi forms, T2 becomes less constricted and less noisy, indicating a loss of glottalization. The duration of T2 is shorter than T44 and T23 in both citation and sandhi forms. 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.

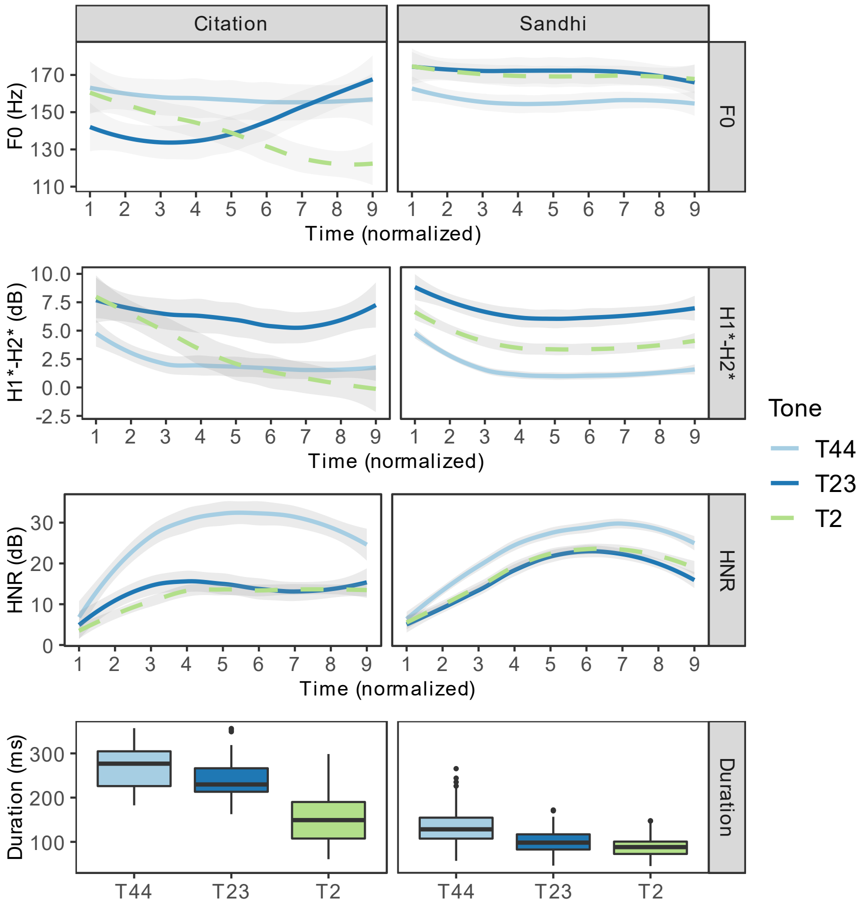

Figure 11. Acoustic parameter values of T44, T23, and T2 in citation and sandhi forms.

3.2.2. Neutralization among T5, T42, and T35

The Sandhi Rule (8) 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 (16).

We compared the three tones in citation forms (T5 vs. T42 vs. T35) in the same model. Figure (12a) shows the LD1 and LD2 distribution of T5, T42, and T35 in citation forms. The classification accuracy of citation forms is 100% and significantly higher than the 45.57% chance level ($p < .001$). We applied the LDA models on every two contrasts of T5, T42, and T35 in their sandhi forms. Figure (12b) shows the LD1 distribution of each tone in each contrast. The classification accuracies were calculated based on leave-one-out cross-validation. The classification accuracies of T35 vs. T42, T5 vs. T35, and T5 vs. T42 in sandhi forms are: 47.06% ($p = .83$; chance = 50.74%), 80.57% ($p < .001$; chance = 52%), and 86.79% ($p < .001$, chance = 53.77%). 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 along these measures, whereas **T5 and T35, and T5 and T42 can still be differentiated significantly above chance**.

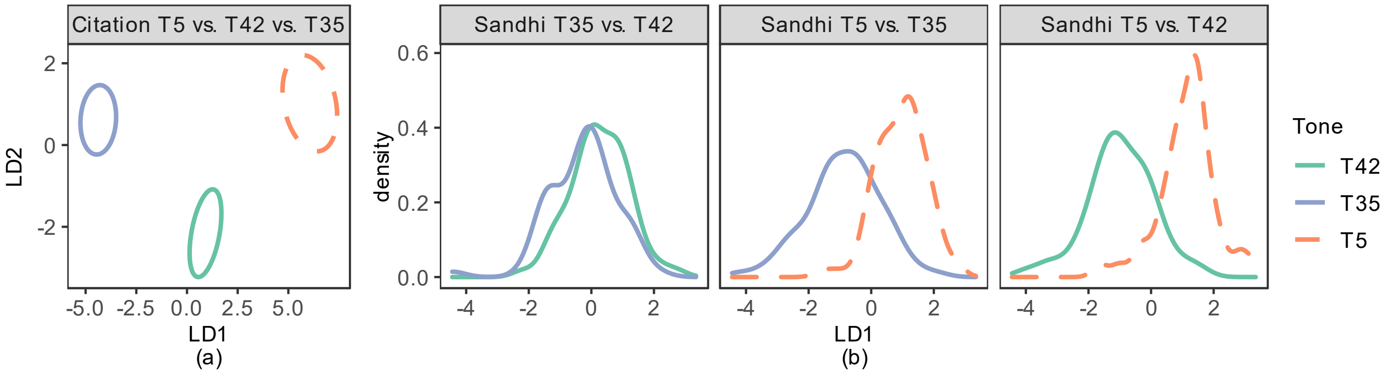

Figure 12. (a) is the LD1-LD2 distribution of T5 vs. T42 vs. T35 in citation forms. The ellipses represent 50% confidence intervals around the mean of each group. (b) is the LD1 distribution of T35 vs. T42, T5 vs. T35, and T5 vs. T42 in sandhi forms, respectively.

We correlate each acoustic parameter with the linear discriminant scores to determine which parameters contribute most to the above-chance discriminations. LD1 explains 92.09% 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 Model (16) and the correlation between all the parameters and the linear discriminant scores are presented in Tables S22-A26 in Supplementary Material 3.

Figure 13 shows values of F0, H1*–H2*, and HNR of T42, T35, and T5 in citation and sandhi forms. In terms of F0, the three tones have well-dispersed contours in citation forms. In sandhi forms, their F0 contours become flat and are largely overlapping. 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, in citation forms, the HNR of T5 has an abrupt fall after Point 4. In sandhi forms, 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 than in citation forms. We compared the H1*–H2* and HNR of T5 between citation and sandhi forms using mixed-effects models, and confirmed both parameters have significantly higher values in sandhi forms than in citation forms. The statistics are in Tables S27-28 in Supplementary Material 3. 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 a loss of glottalization. The duration of T5 is shorter than T42 and T35 in citation and sandhi forms. The duration of T5 is shorter in sandhi forms than in citation forms.

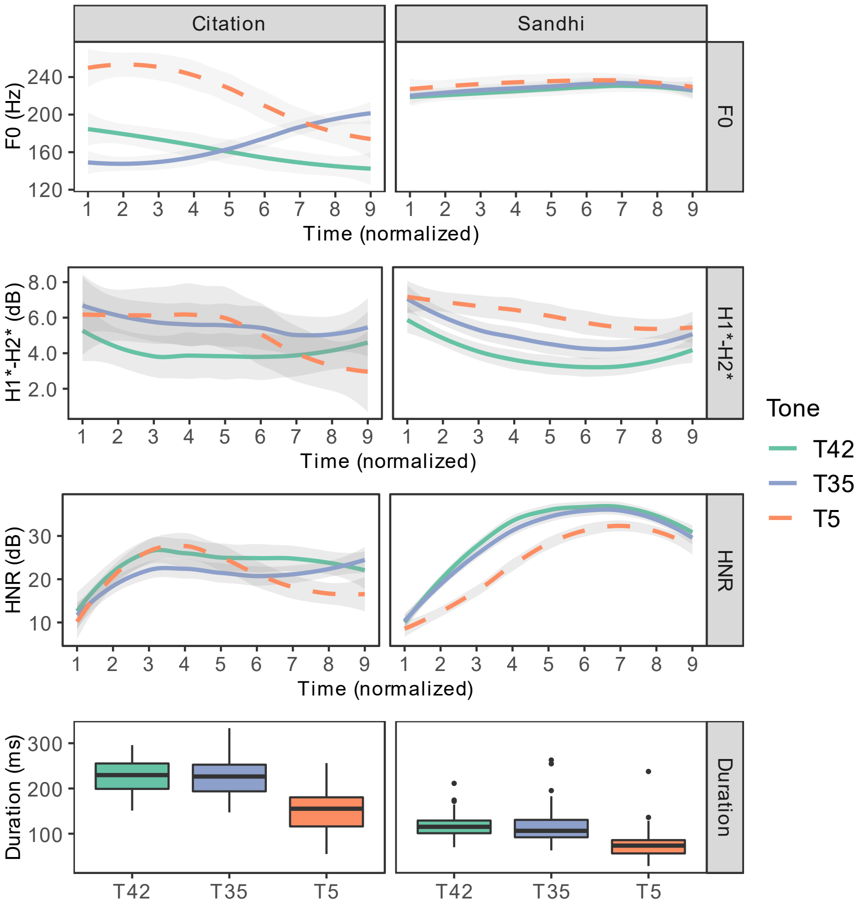

Figure 13. Acoustic parameter values of T42, T35, and T5 in citation and sandhi forms.

Table 11 summarizes 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, four of them are not completely neutralized phonetically: T2-T23, T2-T44, T5-T35, and T5-T42. **All those four 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. Table 12 presents the average duration of each tone in citation and sandhi forms. Checked tones remain to be shorter than unchecked tones in sandhi forms, though the percentage of checked tone duration to unchecked tone duration increases slightly compared with the citation forms (70% vs. 67%).

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 syllables in the stimuli have a voiceless aspirated stop (/t^h/), voiceless affricate (/ts/), or voiceless fricative (/x, θ/) as the onset. Thus, it is possible that the aspirated and fricated onsets introduce noise onto 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. Checked tones in sandhi forms 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 conclude that the glottalized quality of the checked tones 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 in sandhi forms. In sum, after sandhi, duration differentiates checked tones from unchecked tones. Their differences in phonatory quality and F0 are largely neutralized.

Table 11. Classification accuracies and top three parameters that have the highest correlation with LD1.

Citation/Sandhi	Contrast	Classification accuracy (chance level, *p* value)	Parameters Citation/Sandhi
Citation	**T2 vs. T44 vs. T23**	**95% (38.75%, <.001)**	duration, final F0, mid HNR
Sandhi	T23 vs. T44	64.86% (54.05%, .12)	
	T2 vs. T23	**69% (51%, <.001)**	duration, final F0, initial H1*–H2*
	T2 vs. T44	**79.31% (51.72%, <.001)**	duration, initial and final HNR
Citation	**T5 vs. T42 vs. T35**	**100% (45.57%, <.001)**	initial and mid F0, duration
Sandhi	T35 vs. T42	47.06% (50.74%, .83)	
	T5 vs. T35	**80.57% (52%, <.001)**	duration, initial and mid HNR
	T5 vs. T42	**86.79% (53.77%, <.001)**	duration, initial and mid HNR

Table 12. Duration of each tone in citation and sandhi forms (in ms) and the percentage of checked tone duration to unchecked tone duration

	T5	T2	T44	T42	T35	T23	T11	Checked/Unchecked percentage
Citation	**166**	**164**	267	242	229	244	251	67%
Sandhi	**74**	**89**	135	116	114	100	NA	70%

4. Discussion and conclusion

In this study, we have determined how checked tones in Xiapu Min differ from unchecked tones, in terms of their F0 height and contour, phonatory quality, duration, and vowel quality. These parameters are the ones most often associated with checkedness across languages. The results show that in citation forms, checked tones in Xiapu Min differ from unchecked tones in three out of four dimensions. We confirm that the two checked tones – T5 and T2 – have distinct falling contours in comparison with 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 vowels in checked syllables in Xiapu Min are glottalized in the end. Checked tones also have a shorter duration than unchecked tones. However, checked and unchecked tones do not differ in vowel quality. Thus, three out of four 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 four 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. In sandhi forms, 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.

The acoustic results of checked tones in citation and sandhi forms help clarify the relation between short duration and glottalization in Xiapu Min checked syllables. In checked tones in sandhi forms, glottalization is weakened while the short duration is preserved. This suggests that glottalization might serve as a means of achieving a vowel gesture with a shorter duration, which is the intended articulatory target. In citation forms, the short duration of checked syllables is argued to be aided by glottal constriction. Without glottal constriction, and with no sound that follows, voicing would continue until the subglottal pressure drops significantly. But in sandhi forms, the shorter duration can be achieved early by starting the articulatory gesture of the upcoming onset early. It is possible that glottalization might not be needed to ensure a short vowel in sandhi forms.

The results of the acoustic properties of Xiapu Min checked tones in citation and sandhi forms also answer the two inferential questions of this study: which stage of checked syllable sound change is Xiapu Min currently at; what is the next possible stage of checked syllable in the language? As a reminder, we show the three checked syllable sound change paths proposed by Zhu et al. (2008) here again:

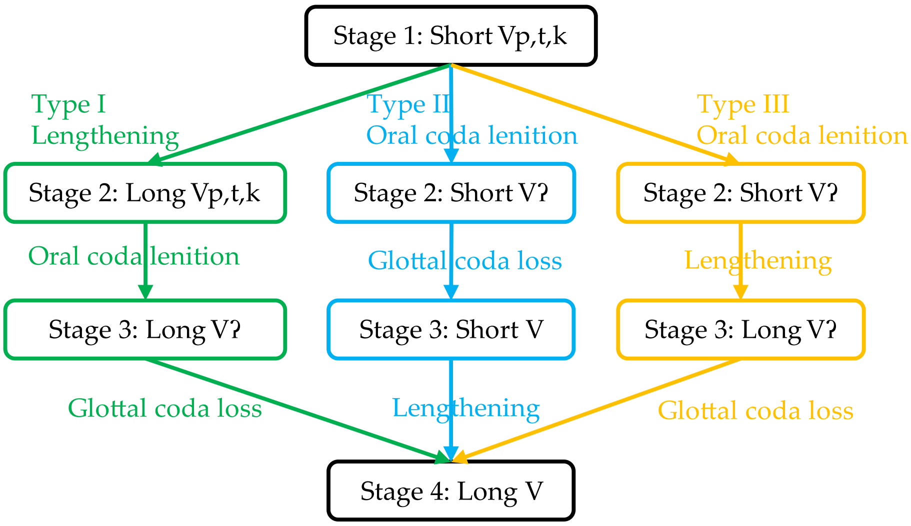

Figure 3. Three paths involving the loss of checkedness (schematized based on Zhu et al. (2008)).

The phonetic features of checked tones in citation forms indicate that Xiapu Min is currently at Stage 2 of either Type II or Type III trajectory because its checked syllables are short and closed by a glottal stop. The observation that glottalization is lost whereas the short duration is retained in sandhi forms suggests that, in Xiapu Min, duration is a more stable feature than glottalization for checked syllables. Thus, the next stage of Xiapu Min checked syllable sound change is more likely to be losing the glottal stop coda than vowel lengthening. Assuming that Xiapu Min checked syllables would go through sound change in the future, its path is most likely to be Type II proposed by Zhu et al. (2008): the glottal stop is lost first, then syllable lengthening takes place.

To conclude: this study has provided the first quantitative acoustic analysis of Xiapu Min tones, revealing the phonetic features of Xiapu Min checked syllables in both citation and sandhi forms. The results provide inference to the diachronic change of Xiapu Min checked syllables and tones. In future work, we plan to conduct perception studies that manipulate F0, phonatory quality, and duration separately in sound signals for both citation and sandhi checked tones. Given that duration appears to be a more stable feature than glottalization in Xiapu Min, we will test whether listeners are more sensitive to duration than glottalization when identifying a tone as checked. And because the duration of checked syllables is significantly shorter than unchecked ones after sandhi neutralization, we will test whether listeners are able to discriminate between checked and unchecked syllables after sandhi based on duration cue.

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.

References

1. Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using Lme4. Journal of Statistical Software 67 (1). https://doi.org/10.18637/jss.v067.i01.
2. Boersma, Paul, and David Weenink. 2021. Praat: Doing Phonetics by Computer [Computer Program] (version 6.2). http://www.praat.org/.
3. Bosch, Nigel, and Luc Paquette. 2018. Metrics for Discrete Student Models: Chance Levels, Comparisons, and Use Cases. Journal of Learning Analytics 5 (2). https://doi.org/10.18608/jla.2018.52.6.
4. Chan, Marjorie K. M. 1987. Tone and Melody in Cantonese. Annual Meeting of the Berkeley Linguistics Society 13 (September): 26. https://doi.org/10.3765/bls.v13i0.1828.
5. Chao, Yuen Ren. 1930. ə sistim əv "Toun-Letəz" [A System of 'Tone-Letters']. Le Maître Phonétique Troisième Série 8 (45): 24–27.
6. Chen, Si, and Caroline R. Wiltshire. 2013. Tone Realization in Younger versus Older Speakers of Nanjing Dialect. In Studies in Chinese Language and Discourse, edited by Zhuo Jing-Schmidt, 2:147–70. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/scld.2.07che.
7. Chien, Yu-Fu, and Allard Jongman. 2019. Tonal Neutralization of Taiwanese Checked and Smooth Syllables: An Acoustic Study. Language and Speech 62 (3): 452–74. https://doi.org/10.1177/0023830918785663.
8. De Maesschalck, R., D. Jouan-Rimbaud, and D.L. Massart. 2000. The Mahalanobis Distance. Chemometrics and Intelligent Laboratory Systems 50 (1): 1–18. https://doi.org/10.1016/S0169-7439(99)00047-7.
9. DiCanio, Christian T. 2009. The Phonetics of Register in Takhian Thong Chong. Journal of the International Phonetic Association 39 (2): 162–88. https://doi.org/10.1017/S0025100309003879.
10. Duan Wenjun and Jia Yuan. 2015. Contrastive Study of Focus Phonetic Realization between Jinan Dialect and Taiyuan Dialect. In 2015 International Conference Oriental COCOSDA Held Jointly with 2015 Conference on Asian Spoken Language Research and Evaluation (O-COCOSDA/CASLRE), 47–52. Shanghai, China: IEEE. https://doi.org/10.1109/ICSDA.2015.7357863.
11. Esposito, Christina M. 2010a. The Effects of Linguistic Experience on the Perception of Phonation. Journal of Phonetics 38 (2): 306–16. https://doi.org/10.1016/j.wocn.2010.02.002.
12. ———. 2010b. Variation in Contrastive Phonation in Santa Ana Del Valle Zapotec. Journal of the International Phonetic Association 40 (2): 181–98. https://doi.org/10.1017/S0025100310000046.
13. ———. 2012. An Acoustic and Electroglottographic Study of White Hmong Tone and Phonation. Journal of Phonetics 40 (3): 466–76. https://doi.org/10.1016/j.wocn.2012.02.007.
14. Garellek, Marc. 2019. The Phonetics of Voice. In Routledge Handbook of Phonetics, edited by William Katz and Peter Assmann, 75–106. Oxford: Routledge.
15. Garellek, Marc, and Christina M. Esposito. 2021. Phonetics of White Hmong Vowel and Tonal Contrasts. Journal of the International Phonetic Association, June, 1–20. https://doi.org/10.1017/S0025100321000104.
16. Garellek, Marc, and Patricia Keating. 2011. The Acoustic Consequences of Phonation and Tone Interactions in Jalapa Mazatec. Journal of the International Phonetic Association 41 (2): 185–205. https://doi.org/10.1017/S0025100311000193.
17. Garellek, Marc, Patricia Keating, Christina M. Esposito, and Jody Kreiman. 2013. Voice Quality and Tone Identification in White Hmong. The Journal of the Acoustical Society of America 133 (2): 1078–89. https://doi.org/10.1121/1.4773259.
18. Gruber, James Frederick. 2011. An Articulatory, Acoustic, and Auditory Study of Burmese Tone. Georgetown University-Graduate School of Arts & Sciences. Ph.D. dissertation, Washington, D.C., USA: Georgetown University. https://repository.library.georgetown.edu/handle/10822/558130.
19. Hall, ROBERT A. Jr. 1971. The Syllable in Italian Phonology. Linguistics 9 (67). https://doi.org/10.1515/ling.1971.9.67.26.

20. Heiberger, Vicki L., and Yoshiyuki Horii. 1982. Jitter and Shimmer in Sustained Phonation. In Speech and Language, 7:299–332. Elsevier. https://doi.org/10.1016/B978-0-12-608607-2.50016-9.
21. Hu, Na, Yuan Jia, and Bin Liu. 2012. Phonetic and Phonological Realization of Narrow Focus in English Declarative Sentences by Zhenjiang EFL Learners. In Proceedings of Speech Prosody 2012, 394–97. Shanghai, China. https://www.isca-speech.org/archive_v0/sp2012/sp12_394.html.
22. Huffman, Marie K. 1987. Measures of Phonation Type in Hmong. The Journal of the Acoustical Society of America 81 (2): 495–504. https://doi.org/10.1121/1.394915.
23. Izenman, Alan Julian. 2013. Linear Discriminant Analysis. In Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning, edited by Alan J. Izenman, 237–80. Springer Texts in Statistics. New York, NY: Springer. https://doi.org/10.1007/978-0-387-78189-1_8.
24. Jia, Xiaoying. 2013. Taiyuan nanjiao xinpai fangyan shengdiao shiyan yanjiu [The experimental study on tones in the southern suburb's new-style dialect in Taiyuan]. Master's Thesis, Shanxi, China: Shanxi University. https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=%E5%A4%AA%E5%8E%9F%E5%8D%97%E9%83%8A%E6%96%B0%E6%B4%BE%E6%96%B9%E8%A8%80%E5%A3%B0%E8%B0%83%E5%AE%9E%E9%AA%8C%E7%A0%94%E7%A9%B6&btnG=.
25. Jia, Yuan, and Aijun Li. 2012. Phonetic Realization of Accent from Chinese English Learners in Various Dialectal Regions. In 2012 8th International Symposium on Chinese Spoken Language Processing, 296–300. Kowloon Tong, China: IEEE. https://doi.org/10.1109/ISCSLP.2012.6423547.
26. Keating, Patricia, Marc Garellek, and Jody Kreiman. 2015. Acoustic Properties of Different Kinds of Creaky Voice. In Proceedings of the 18th International Congress of Phonetic Sciences. Glasgow.
27. Keene, Oliver N. 1995. The Log Transformation Is Special. Statistics in Medicine 14 (8): 811–19. https://doi.org/10.1002/sim.4780140810.
28. Klatt, Dennis H., and Laura C. Klatt. 1990. Analysis, Synthesis, and Perception of Voice Quality Variations among Female and Male Talkers. Journal of the Acoustical Society of America 87: 820–57.
29. 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): EL509–15. https://doi.org/10.1121/1.5043094.
30. Kuo, Chen-Hsiu. 2013. Perception and Acoustic Correlates of the Taiwanese Tone Sandhi Group. Los Angeles: University of California, Los Angeles. https://escholarship.org/uc/item/30q6w11t#main.
31. Lai, Wenpan. 2016. Jiyu Xiamen Hua Zuo Minnan Fangyan Shengdiao Yanjiu [Acoustic Analysis of Tones in Southern Min Dialect Based on Xiamen Dialect]. Art Science and Technology, no. 4: 129–30.
32. Lalhminghlui, Wendy, and Priyankoo Sarmah. 2018. Production and Perception of Rising Tone Sandhi in Mizo. In 6th International Symposium on Tonal Aspects of Languages (TAL 2018), 114–18. ISCA. https://doi.org/10.21437/TAL.2018-23.
33. Li, Bing, and Yanni Liu. 2006. Changsha fangyan danzidiao ji biandiao de shiyan yuyinxue baogao [An acoustic of study of citation tones and tone sandhi in Changsha Chinese]. Journal of Hunan University (Social Sciences) 20 (4): 107–12.
34. Li, Xinghui. 2004. Gu rushengzi zai xiangyu zhongde fenhua [Division of the entering tone words in Xiang dialect]. Journal of Central South University (Social Science) 10 (3): 394–97.
35. Li, Yang. 2016. Complete and Incomplete Neutralisation in Fuzhou Tone Sandhi. In 5th International Symposium on Tonal Aspects of Languages (TAL 2016), 116–20. ISCA. https://doi.org/10.21437/TAL.2016-25.
36. Lin, Hwei-Bing, and Bruno H. Repp. 1989. Cues to the Perception of Taiwanese Tones. Language and Speech 32 (1): 25–44. https://doi.org/10.1177/002383098903200102.
37. Liu, Lili. 2008. Jipin guiyi he diaoxi guizheng de fangyan shiyan [F0 normalization and tone adjustment in dialect experiments]. Chinese Journal of Phonetics, no. 1: 221–27.
38. Liu, Zhangcai. 2013. Hunan Yuanjiang Chishan hua danzidiao he shuangzidiao shengxue shiyan yanjiu [An acoustic study of mono-syllable tone and tone sandhi of Yuanjiang Chishan dialect in Hu-nan province]. Master's Thesis, Guangxi, China: Guangxi Normal University. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CMFD&dbname=CMFD201402&filename=1013245249.nh&uniplatform=NZKPT&v=5HIyN8FSwIE3J8WTNeVYPZZ6IIPpXy1SfH1OvY8GeADP9TrbCL5u2Ga5Pi%25mmd2FiNpmS.
39. Oakden, Christopher. 2017. Checked Tone Merger in the Nanjing Dialect: An Acoustic Analysis. In Proceedings of the 29th North American Conference on Chinese Linguistics (NACCL-29), edited by Lan Zhang, 1:141–52. Columbus, OH. https://naccl.osu.edu/sites/default/files/Cover_Vol-1-naccl29.pdf.
40. Pan, Ho-hsien. 2005. Voice Quality of Falling Tones in Taiwan Min. In Proceedings of INTERSPEECH 2005, 1401–4. Lisbon, Portugal. https://www.isca-speech.org/archive/interspeech_2005/i05_1401.html.
41. ———. 2017. Glottalization of Taiwan Min Checked Tones. Journal of the International Phonetic Association 47 (1): 37–63. https://doi.org/10.1017/S0025100316000281.
42. Pan, Ho-hsien, Hsiao-tung Huang, and Shao-ren Lyu. 2016. Coda Stop and Taiwan Min Checked Tone Sound Changes. In Interspeech 2016, 1011–15. ISCA. https://doi.org/10.21437/Interspeech.2016-597.
43. Pan, Ho-hsien, and Shao-ren Lyu. 2021. Taiwan Min Nan (Taiwanese) Checked Tones Sound Change. In Interspeech 2021, 2641–45. ISCA. https://doi.org/10.21437/Interspeech.2021-672.

44. Peirce, Jonathan, Jeremy R. Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman, and Jonas Kristoffer Lindeløv. 2019. PsychoPy2: Experiments in Behavior Made Easy. Behavior Research Methods 51 (1): 195–203. https://doi.org/10.3758/s13428-018-01193-y.
45. Pickett, Velma B., María Villalobos Villalobos, and Stephen A. Marlett. 2010. Isthmus (Juchitán) Zapotec. Journal of the International Phonetic Association 40 (3): 365–72. https://doi.org/10.1017/S0025100310000174.
46. Qin, Zhen, and Peggy P. K. Mok. 2014. Discrimination of Cantonese Tones by Speakers of Tone and Non-Tone Languages. Kansas Working Papers in Linguistics, January. https://doi.org/10.17161/KWPL.1808.12864.
47. Ramsey, Jim, and Brian Ripley. 2017. Pspline: Penalized Smoothing Splines (version 1.0-18). https://CRAN.R-project.org/package=pspline.
48. Ratliff, Martha Susan. 2010. Meaningful Tone: A Study of Tonal Morphology in Compounds, Form Classes, and Expressive Phrases in White Hmong. DeKalb: Northern Illinois University Press.
49. Seyfarth, Scott, and Marc Garellek. 2018. Plosive Voicing Acoustics and Voice Quality in Yerevan Armenian. Journal of Phonetics 71 (November): 425–50. https://doi.org/10.1016/j.wocn.2018.09.001.
50. Shao, Dandan. 2012. Jiyu EGG de Meixian, Fuzhou, Changsha Fangyan Shengdiao Shiyan Yanjiu [Acoustic Experimental Analysis of Meixian, Fuzhou, and Changsha Dialect Based on EGG Data]. Master's Thesis, Nanjing, China: Nanjing Normal University. https://cdmd.cnki.com.cn/Article/CDMD-10319-1013105547.htm.
51. Shi, Feng, Qibin Ran, and Ping Wang. 2010. On Sound Pattern [Lun Yuyin Geju]. Nankai Linguistics, no. 1: 1–14.
52. Shue, Yen-Liang, Patricia A. Keating, Chad Vicenik, and Kristine Yu. 2011. VoiceSauce: A Program for Voice Analysis. In Proceedings of the International Congress of Phonetic Sciences, 1846–49. Hong Kong.
53. Speck, Charles H. 1978. The Phonology of Texmelucan Zapotec Verb Irregularity. Master's Thesis, Grand Forks, ND: University of North Dakota. https://commons.und.edu/theses/2660.
54. Su, Hui. 2016. Jiangdu fangyan danzidiao shiyan yanjiu [Experimental analysis of Jiangdu Dialect citation tone]. Modern Chinese, no. 4: 19–23.
55. Sun, Huaxian. 2003. Nanjing Fangyan Shengdiao de Ruogan Wenti [A Discussion on the Tones of Nanjing Dialect]. Journal of Nanjing Xiaozhuang College, 34-40, 19 (1). http://www.cqvip.com/qk/85356x/200301/1001331933.html.
56. Tang, Zhiqiang. 2014. Wanshu Jianghuai guanhua rusheng shiyan yanjiu [Experimental analysis of checked tone in Anhui Jianghuai Mandarin]. Master's Thesis, Nanjing, China: Nanjing Normal University. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CMFD&dbname=CMFD201501&filename=1014347991.nh&uniplatform=NZKPT&v=iP-wMUEH7qe0ikdrD9kZAlc%25mmd2F2xJSf61Ck42RSLQm0obCKMmGjvuBULmEjVjJJHQb1.
57. Teodocio Olivares, Amador. 2009. Betaza Zapotec Phonology : Segmental and Suprasegmental Features. Master's Thesis, Austin, TX: University of Texas at Austin. https://repositories.lib.utexas.edu/handle/2152/19162.
58. Trask, R. L. 1996. A Dictionary of Phonetics and Phonology. https://www.taylorfrancis.com/books/9781134831012.
59. Venables, W. N., Brian D. Ripley, and W. N. Venables. 2002. Modern Applied Statistics with S. 4th ed. Statistics and Computing. New York: Springer.
60. Wen, Jing. 2015. Mindongqu Xiapu fangyan yuyin yanjiu [A study on the phonology of Xiapu dialect]. Master's Thesis, Fujian, China: Fujian Normal University. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CMFD&dbname=CMFD201601&filename=1015720187.nh&uniplatform=NZKPT&v=taJQvNmlO3IpPKCgVcRTtW2y8H1jQJifrIuy-GsKi9UELK4n6ZFJ2i5dENKvLXYqr.
61. Wu, Bo. 2018. An Acoustic Analysis of Vowels in Checked Syllables in Chinese. Chinese Journal of Acoustics, no. 04: 491–502. https://doi.org/10.15949/j.cnki.0217-9776.2018.04.009.
62. Xia, Liping, and Fang Hu. 2016. Vowels and Diphthongs in the Taiyuan Jin Chinese Dialect. In Interspeech 2016, 993–97. ISCA. https://doi.org/10.21437/Interspeech.2016-249.
63. Xiapu Government. 2021. Xiapu Xian Diqici Quanguo Renkou Pucha Gongbao [Xiapu County Seventh National Census Report]. http://www.xiapu.gov.cn/zwgk/zfxxgkzdgz/tjxx/tjgb/202106/t20210628_1491098.htm.
64. Yang, Yongkai, and Ying Chen. 2018. Effects of Entering Tone on Vowel Duration and Formants in Nanjing Dialect. In Studies on Speech Production, edited by Qiang Fang, Jianwu Dang, Pascal Perrier, Jianguo Wei, Longbiao Wang, and Nan Yan, 10733:146–57. Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-00126-1_14.
65. Zhu, Xiaonong. 2004. Jipin Guiyihua - Ruhe Chuli Shengdiao de Suiji Chayi? [F0 Normalization: How to Deal with Between Speaker Tonal Variations?]. Linguistic Sciences 3 (2): 3–19.
66. Zhu, Xiaonong, Lei Jiao, Zhicheng Yan, and Ying Hong. 2008. Rusheng yanhua santu [Three ways of Rusheng sound change]. Studies of the Chinese Language, no. 4: 324–38.