

Article

Checked Syllables, Checked Tones, and Tone Sandhi in Xiapu Min

Yuan Chai ^{1,*}  and Shihong Ye ²¹ Department of Linguistics, University of California San Diego, La Jolla, CA 92093, USA² Second High School Attached to Beijing Normal University, Beijing 100088, China; yeshihong@bnu.edu.cn

* Correspondence: yuc521@ucsd.edu

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 investigated the acoustic features of checked syllables in citation and sandhi forms in Xiapu Min, an under-documented language from China. We conducted a production experiment and analyzed the F₀, phonatory quality, vowel duration, and vowel quality 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 become less glottalized. However, the shorter 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 become less glottalized 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; tone sandhi; phonetic neutralization



Citation: Chai, Yuan, and Shihong Ye. 2022. Checked Syllables, Checked Tones, and Tone Sandhi in Xiapu Min. *Languages* 7: 47. <https://doi.org/10.3390/languages7010047>

Academic Editors: Christina Esposito and Sameer Khan

Received: 6 May 2021

Accepted: 11 February 2022

Published: 24 February 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Xiapu (霞浦) Min is a variety of Eastern Min spoken in Xiapu County (Ningde, Fujian), China (see Figure 1 for the map). There have been 475,936 residents living in Xiapu County in 2021 ([Xiapu County Bureau of Statistics 2021](#)). Checked syllables in Xiapu Min are syllables that are closed by a 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; and to determine whether the contrast between checked vs. unchecked tones is neutralized in sandhi forms.

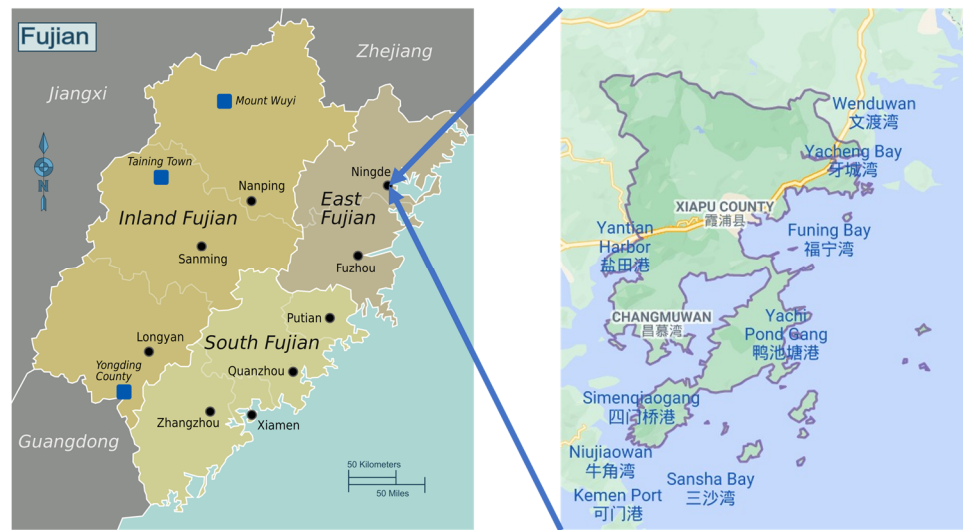


Figure 1. Map of Xiapu County. (retrieved from https://en.wikivoyage.org/wiki/File:Fujian_map.png accessed on 14 February 2022 and <https://www.google.com/maps> accessed on 14 February 2022).

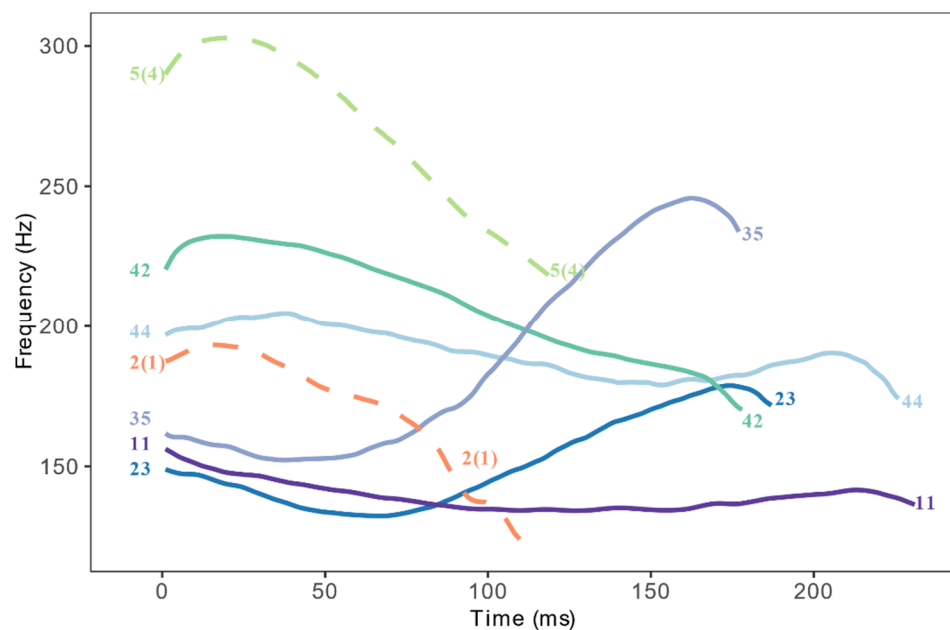


Figure 2. F0 track of /θi/ in seven tones by a female speaker.

The term “checked” has heterogeneous definitions among different languages of different families. We focus on checked syllables that are closed with an oral or a glottal stop and that also form a prosodic opposition against unchecked syllables. In Xiapu Min, two lexical tones are restricted to V? syllables, whereas the other five lexical tones are restricted to V and VN syllables. This is the evidence that V? syllables are in prosodic opposition to V and VN syllables. Note that we consider that Xiapu Min has V? checked syllables associated with checked tones, rather than having glottalized tones realized on V syllables, because there is phonological evidence that the glottal stop is a segment in the language. In Xiapu Min, V? syllables contrast with V and VN syllables in onset change pattern in disyllabic compounds. The phonological rules and examples are presented in Table 1. In a disyllabic compound, the /t/ onset of the second syllable becomes [r] when it follows a V syllable (a), [n] following a VN syllable (b), and remains [t] when it follows a V? syllable (c). This phenomenon has also been reported in Wen (2015). Table 1 Examples (a)–(c) demonstrate a contrast among V, VN, and V? in phonological transformations, indicating

that V? is a contrastive syllable type in Xiapu Min. With the evidence of tone restriction and onset change pattern, V? in Xiapu Min fits our definition of “checked syllable” that will be used in this paper. Such checked syllables are widely reported in Chinese languages: Taiwanese Min (Kuo 2013; Pan 2017); Yun’ao Min (Zhang 2020); Xiamen Min (Lai 2016); Shanghai Wu (Zee and Maddieson 1979); Nanjing Jianghuai Mandarin (Sun 2003; Yang and Chen 2018); and Meixian Hakka (Shao 2012).

Table 1. Onset change pattern after different types of syllables in Xiapu Min *.

	Phonological Rules	Examples
(a)	/t/ → [t]/CV _σ + σ___	/t ^h e 42 tain 23/ → [t ^h e 55 rain 23] 体重 “body weight”
(b)	/t/ → [n]/CVN _σ + σ___	/poŋ 44 to? 5/ → [poŋ 44 no? 5] 饭桌 “dining table”
(c)	/t/ → [t]/CV? _σ + σ___	/t ^h e? 5 to 23/ → [t ^h e 55 to 23] 铁路 “railroad”

* Note that the phenomenon in Table 1 is undergoing change 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.

In other languages, “checkedness” is not associated with syllables exclusively. It can also be a type of phonation associated with vowels. The status of the glottal stop determines whether checkedness is a phonation or a syllable type. For example, the glottal stop in V? 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 V? syllable (d) behave the same as nouns with a V syllable (e), but differently from nouns with a VC syllable (f) (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
(d)	/ju?/	“house”	[ju?m]	“her house”
(e)	/lo/	“face”	[lom]	“her face”
(f)	/ʃab/ *	“clothes”	[ʃab mi] *	“her clothes”

* Speck (1978) transcribed “clothes” as /ʃab/. Symbol /ʃ/ is not an IPA symbol. We replace /ʃ/ in Speck (1978) with /j/ here based on convention and the phonological description of /ʃ/ in Speck (1978).

In order to capture the phonetic features of checkedness to the largest extent, this paper reviews the acoustic correlates of all kinds of purported checked syllables and vowels in the literature. Whether checked syllables and checked vowels can differ in their phonetic realization is beyond the scope of the current paper and will not be further discussed. 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 under-documented and under-studied language, will in return help clarify the phonetic nature of the checked syllables and vowels in general. Four parameters have been found to distinguish checked syllables and vowels from unchecked ones: **F0**, **duration**, **voice quality**, and **vowel quality**. In Section 2, we review how checked and unchecked syllables and vowels differ in each of those four parameters in the literature.

2. Acoustic Correlates of Checked Syllables and Vowels

2.1. F0

Many Chinese languages have checked tones associated with checked syllables, which have a distinct F0 from unchecked tones. We surveyed eleven varieties of Chinese languages that have distinct F0 for checked tones from unchecked ones. The tonal values were supported by the F0 measurement of tones in the references. The languages include: Nanjing (Chen and Wiltshire 2013; Yang and Chen 2018), Anqing, Wuhu, and East Hefei (Tang 2014) varieties of Jianghuai Mandarin; Meixian Hakka (Shao 2012); Chishan (Liu 2013) and Changsha (Li and Liu 2006; Shao 2012) varieties of Xiang; Xiamen (Lai 2016), Taiwanese (Kuo 2013; Pan 2017), and Fuzhou (Shao 2012) varieties of Min; and South

Taiyuan Jin (Jia 2013). A summary of the tone system of each variety is presented in Table S1 in Supplementary Material S3.

Apart from Chinese languages, there are other languages that were reported to have checked tones. In Burmese, the checked tone is associated with checked syllables closed by a 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 a “creaky” tone and analyzed the glottalization at the end of the vowel as a suprasegmental creaky phonation (V̤). 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 and Mok 2014). Two out of five tones in Nantong Chinese can be carried by both checked and unchecked syllables (Ao 1993). The checked phonation in San Melchor Betaza Zapotec (Teodocio Olivares 2009) and Isthmus Zapotec (Pickett et al. 2010) is not tone-dependent. Checked and unchecked vowels can be associated with the same set of tones in these languages.

2.2. Duration

Checked syllables and vowels have been found to be shorter than unchecked ones in a number of languages: Nanjing (Yang and Chen 2018; Wu 2018), Hefei, Nantong (Ao 1993; Wu 2018), Anqing, and Wuhu (Tang 2014) Jianghuai Mandarin; Meixian Hakka (Shao 2012); Fuzhou (Shao 2012), Xiamen (Lai 2016), and Taiwanese Min (Lin and Repp 1989; Kuo 2013); South Taiyuan Jin (Jia 2013); Burmese (Gruber 2011). The White Hmong checked/creaky tone has a shorter duration than other tones in general (Esposito 2012), but the duration is dependent on vowel quality (Garellek and Esposito 2021). In some languages, there is a length difference between their two checked tones. In Meixian Hakka, checked T5 is shorter than checked T32. In Fuzhou Min, checked T5 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 shorter duration. Checked tones in Changsha Xiang (Shao 2012), Chishan Xiang (Liu 2013), and Anqing Jianghuai Mandarin (Tang 2014) have a longer duration than unchecked tones. Hong Kong Cantonese has both short checked and long checked tones. The long checked tone has a 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 short modal vowels in open syllables (Teodocio Olivares 2009).

2.3. Quality of Phonation

Several studies reported that checked tones were realized with non-modal phonations. Differences in phonatory quality have been represented by the articulatory measurements of glottal opening (e.g., Open Quotient measured from electroglottography: Esposito 2012; Gruber 2011; Pan 2017; Pan et al. 2016; Shao 2012; Tang 2014) or the acoustic correlates of glottal constriction and voicing periodicity (e.g., spectral tilt, HNR, CPP, and jitter and shimmer) (Esposito 2012; Garellek and Esposito 2021; Gruber 2011; Kuo 2013; Pan 2017; Pan et al. 2016; Shao 2012). Shao (2012) found that checked T5 in Meixian Hakka and Fuzhou Min was more glottalized than unchecked tones, whereas checked T32 in Meixian Hakka and checked T24 in Fuzhou Min were realized with a breathy voice. Checked T5 in East Hefei Jianghuai Mandarin was more glottalized than unchecked ones, especially at the end of vowels (Tang 2014). Burmese checked tone (Gruber 2011) and White Hmong

checked/creaky tone (Esposito 2012; Garellek and Esposito 2021) were reported to have a more glottalized quality than unchecked tones based on acoustic measurements.

There are also checked syllables and vowels that are not associated with a glottalized quality. Checked syllables in Nanjing Jianghuai Mandarin are found to be frequently produced without a glottal coda. Vowels in checked syllables have lower jitter values than those in unchecked syllables, suggesting that checked syllables are more modal than unchecked ones (Oakden 2017). Checked tones in Taiwanese Min are not consistently glottalized (Pan 2005, 2017). The /p, t, k, ʔ/ codas in checked syllables are frequently deleted in natural production, resulting in vowel lengthening and an increase in voicing periodicity (Pan et al. 2016; Pan and Lyu 2021). Checked tones in Changsha Xiang (Shao 2012) and Anqing and Wuhu Jianghuai Mandarin (Tang 2014) have a similar degree of glottal opening as unchecked tones, indicating a modal quality. Finally, for Vietnamese, vowels with checked tones (D1: rising checked; D2: low checked) have less glottalization than vowels in unchecked syllables carrying unchecked tones (B1: rising unchecked; B2: low glottalized unchecked) (Michaud 2004).

2.4. Vowel Quality

Yang and Chen (2018) report that for Nanjing Jianghuai Mandarin, compared with vowels in unchecked syllables, vowels in checked syllables have 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 observe that the vowel quality difference is due to the glottal constriction gesture at the end of the vowel. The glottal constriction causes jaw lowering and consequently a lower and fronter tongue position, resulting in formant value raising. Wu (2018) measured the vowel space in Nanjing, Nantong, and Hefei Jianghuai Mandarin, and found that the F1 of vowels in checked syllables is higher than that of unchecked ones. Front and back vowels are more concentrated in the middle position on the F2 scale. In Taiyuan Jin, the number of vowel contrasts is reduced in checked syllables compared with unchecked syllables (Xia and Hu 2016). While V and VN syllables allow six and five vowel contrasts, respectively, Vʔ syllables only allow two central vowels /ɐ, ə/.

In Table 3, we summarize the phonetic features of the checked constituents in all the languages mentioned in Sections 2.1–2.4. 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.

Table 3. Phonetic features of selected languages with checked constituents (Y for “Yes”; N for “No”; NA if a feature was not described in the reference.).

	Distinct Pitch	Shorter Duration	Non-Modal Voice Quality	Different Vowel Quality	Sources
Nanjing Jianghuai Mandarin	Y	Y	N	Y	(Chen and Wiltshire 2013; Yang and Chen 2018; Oakden 2017; Wu 2018)
Meixian Hakka	Y	Y	Y	NA	(Shao 2012)
Changsha Xiang	Y	N	N	NA	(Li and Liu 2006; Shao 2012)
Chishan Xiang	Y	N	NA	NA	(Liu 2013)
Xiamen Min	Y	Y	NA	NA	(Lai 2016)
Anqing Jianghuai Mandarin	Y	N	N	NA	(Tang 2014)
Wuhu Jianghuai Mandarin	Y	Y	N	NA	(Tang 2014)
East Hefei Jianghuai Mandarin	Y	Y	Y	Y	(Tang 2014; Wu 2018)
Taiwanese Min	Y	Y	N	NA	(Kuo 2013; Pan 2005, 2017; Pan et al. 2016; Pan and Lyu 2021)
South Taiyuan Jin	Y	Y	NA	Y	(Jia 2013; Xia and Hu 2016)
Fuzhou Min	Y	Y	Y	NA	(Shao 2012)
Nantong Jianghuai Mandarin	N	Y	NA	Y	(Ao 1993; Wu 2018)
Cantonese	N	Y/N *	NA	NA	(Chan 1987; Qin and Mok 2014; Zhu et al. 2008)
Burmese	Y	Y	Y	NA	(Gruber 2011)
White Hmong	Y	Y	Y	NA	(Esposito 2012; Garellek and Esposito 2021)
San Melchor Betaza Zapotec	N	N	Y	NA	(Teodocio Olivares 2009)
Isthmus Zapotec	N	NA	Y	NA	(Pickett et al. 2010)

* Cantonese has both short and long checked tones.

2.5. Acoustic Correlates of Checked Tones in Sandhi Forms

Compared to the phonetics of checked tones in citation forms, the phonetic properties of checked tones in sandhi forms are less studied. Pan (2017) studied the voice quality of checked tones in sandhi in Taiwanese Min. In Taiwanese Min, a tone undergoes sandhi when it is in the middle of a phrase but preserves its citation form in phrase-final position. The low checked tone /3/ is realized as [5] after sandhi, while the high checked tone /5/ is realized as [3]. Pan (2017) found that 80% of /ʔ/ codas in checked tones are deleted in both sandhi and citation forms. Codas of /p, t, k/ are more frequently lenited to [ʔ] in sandhi forms than in citation forms. The high checked tone [5] is produced with less vocal fold contact than the unchecked high-falling tone [51] in both citation and sandhi forms. The low checked tone [3] is produced with less vocal fold contact but a noisier quality than the unchecked low-falling tone [31] in sandhi forms (Pan 2017). Chien and Jongman (2019) compared sandhi checked tones with citation checked tones that are phonologically neutralized (i.e., [3] /5/ vs. [3] /3/; [5] /3/ vs. [5] /5/) in Taiwanese Min. They found that the F0 height and contour of sandhi checked tones and citation checked tones are indistinguishable after neutralization.

Checked tones in Xiapu Min also undergo tone sandhi in specific environments. While Pan (2017) focused on the voice quality of sandhi checked tones, and Chien and Jongman (2019) focused on the F0 of sandhi checked tones, we will investigate three features that have been found to be related to checked syllables and vowels—F0, duration, and phonatory quality. Another significance of studying the phonetic feature of checked tones in sandhi forms in Xiapu Min is that it can demonstrate the phonetic consequences when checked tones change into unchecked ones. Taiwanese Min checked tones remain phonologically checked after sandhi. In contrast, Xiapu Min checked tones acquire the same F0 target as unchecked tones after sandhi, indicating that checked and unchecked tones and syllables are possibly neutralized. Table 4 lists the relevant tone sandhi rules. There are two checked tones in Xiapu Min: high-falling T5 and low-falling T2. Tone sandhi occurs when two tones are juxtaposed. Low-falling checked T2 becomes mid-level unchecked T44 (Rule g), whereas high-falling checked T5 becomes high-level unchecked T55 (Rule h) when they are followed by another tone in compounds.

Table 4. 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 fieldwork data collected by the authors).

Phonological Rules	
(g)	/T2, T23, T44/ → [T44] / ___ X
(h)	/T5, T35, T42/ → [T55] / ___ X

Rule (g) in Table 4 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; Fuzhou Min T232-T44 neutralization: Li 2016) or complete (e.g., Taiwanese Min: Chien and Jongman 2019; Fuzhou Min T53-T44 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 (g) in Table 4, the citation checked T2 should be realized as unchecked T44 before another tone. However, 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.

In summary, 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 completely neutralized in terms of their F0, phonation type and duration? To answer the first question, we will comprehensively measure the acoustic properties 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 to determine which acoustic parameters can effectively differentiate those neutralized tones.

3. Materials and Methods

3.1. Stimuli

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

The reading material for the production experiment consisted of two parts. In Part 1, we asked the participants to produce minimal pairs of citation tones. The stimuli of Part 1 had 90 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.

In Part 2, we 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 the effects of the onset on the preceding vowel. Table 5 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 presented in Supplementary Material S1. 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 5. 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	/xu? 2 tsɔŋ 44/ → [xu 44 tsɔŋ 44] /xu 23 kain 44/ → [xu 44 kain44]	服装 “clothes” 护工 “caregiver”
	T23 vs. T44	/to 23 k ^h eu 42/ → [to 44 k ^h eu 42] /to 44 k ^h eu 42/ → [to 44 k ^h eu 42]	路口 “intersection” 刀口 “wound”
{T5, T35, T42} → T55 / ___ X	T2 vs. T44	/tsa? 2 ki 44/ → [tsa 44 ki 44] /tsa 44 kaŋ 44/ → [tsa 44 kaŋ 44]	杂技 “acrobatics” 查岗 “check up”
	T5 vs. T35	/θi? 5 tɕin 42/ → [θi 55 tɕin 42] /θi 35 tɕia 42/ → [θi 55 tɕia 42]	湿疹 “eczema” 试纸 “test paper”
	T35 vs. T42	/ka 35 kai 42/ → [ka 55 kai 42] /ka 42 θe 42/ → [ka 55 θe 42]	价格 “price” 假设 “hypothesize”
	T5 vs. T42	/ka? 5 pain 42/ → [ka 55 pain 42] /ka 42 tsein 42/ → [ka 55 tsein 42]	甲板 “deck” 假钱 “fake money”

All stimuli were presented in Chinese characters on a computer screen in 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 Realtek ALC236 soundcard.

3.2. Acoustic Measures in Use and Criteria for Detecting Tracking Errors and Outliers

We segmented the vowel for each target syllable. All target syllables are CV or CV?. 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 and Weenink 2021). The formant setting was to find 5 formants in the 0–5000 Hz range. H1*–H2* and HNR together represent the voice quality of the target vowels. 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 will trigger 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 and 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 from the analyses of all acoustic measures. Next, we performed visual inspections of 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 V? token. We drew an F0 track for every V? 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 the correct estimation of F0. The pitch track plots and the excluded tokens are presented in Figure S1 and Table S2 in Supplementary Material S3.

We also performed visual and statistical inspection of the formant outputs to exclude tracking errors from formant analysis. H1*–H2* was 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 presented in Figures S2 and S3 and Table S3 in Supplementary Material S3.

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 of 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 were no longer balanced by participant or tone. There were more tracking errors for the vowel /u/ than for other vowels. After excluding tracking errors for formants, F1 and F2 became unbalanced by participant and vowel. Using a z-score by participant for H1*–H2*, F1, and F2 could therefore distort the data and obscure the effects. The R code for detecting tracking errors and outliers and all the statistical analyses in the following sections are presented in Supplementary Material S4.

4. Results

4.1. 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 are 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 that 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 720 tokens in total (90 segments * eight participants). A total of 91 tokens were excluded because of either corrupted recording or mispronunciations, leaving 629 tokens valid for analysis. VoiceSauce ([Shue et al. 2011](#)) yielded 137,226 data points in total. Based in part on feedback from the participants, we believe there are three reasons for the large number of mispronunciations. First, some of the tokens were uttered in Mandarin due to the influence of Chinese characters presented in the elicitation materials. Second, some target syllables seldom occur in isolation in the language, but co-occur with other syllables as a compound. The participants sometimes pronounced the target syllable with its sandhi tone in a compound instead of with a citation tone in isolation. Third, each target syllable was only elicited once for each participant, and the unfamiliarity with the elicitation material might have led to some pronunciation errors. The tracking error and outlier detection and exclusion procedures were the same as described in Section 3.2. After data exclusion, there were 133,246, 125,700, 125,700, 119,838, and 136,469 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 equally timed intervals, and the mean of each interval was calculated. The descriptive statistics of the dataset can be found in Tables S4–S7 in Supplementary Material S3.

4.1.1. F0

Figure 3 shows the average F0 value of each tone over nine equally timed intervals for female and male participants, respectively. Checked T2 and 5 are represented by dotted lines. For both female and male speakers, checked T5 has the highest F0 among all tones. Checked T2 has a similar onset as the rising T35. In general, the tones produced by female speakers have higher F0 values than those produced by male speakers. T44, 23, 2, 11 are in a more compressed F0 range for male speakers than female speakers.

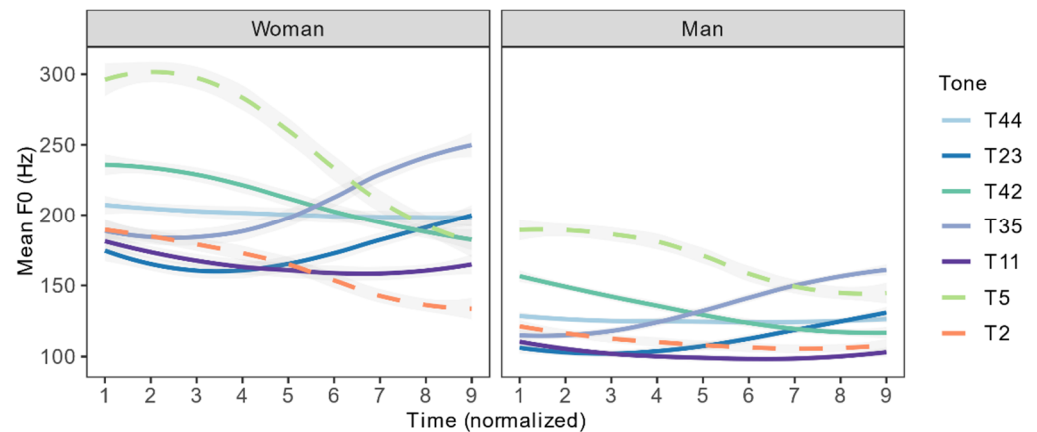


Figure 3. Average F0 track for female and male speakers of the seven tones in Xiapu Min.

Figure 3 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 it has been used in several studies (Duan and Jia 2015; Hu et al. 2012; Jia and Li 2012). The calculation of the log z-score was described in Section 3.2. 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 4a.

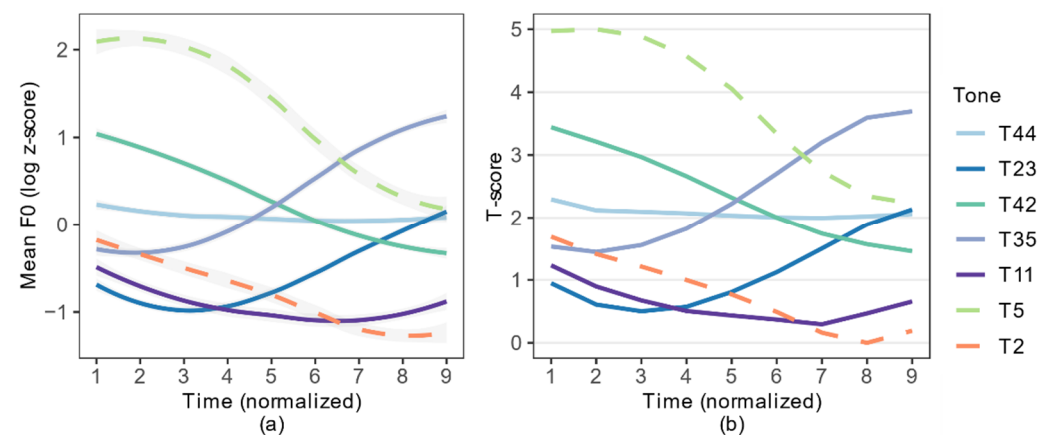


Figure 4. Log z-score (a) and T-score (b) of F0 over nine equally timed intervals of seven tones in Xiapu Min.

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. The T-score is calculated using Formula (1). F0 represents the F0 value of the current time point. F0_{min} and F0_{max} are the minimum and maximum values of F0 among all time points. The correspondence between T-score and Chao numeral is presented in Table 6. 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.

$$T = \frac{\log_{10}(F0) - \log_{10}(F0_{\min})}{\log_{10}(F0_{\max}) - \log_{10}(F0_{\min})} \times 5 \tag{1}$$

Table 6. T-score and Chao numeral conversion (Liu 2008).

T-Score	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 (1) into Formula (2) for calculating the T-score. Rather than using the logged value of F0, we used 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 the log z-score in Figure 4a, but it 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 equally timed intervals are shown in Figure 4b.

$$T = \frac{LZ(F0) - LZ(F0_{\min})}{LZ(F0_{\max}) - LZ(F0_{\min})} \times 5 \tag{2}$$

The T-score at Time Point 1 and 9 of each tone, and their corresponding Chao numeral based on Table 6, are listed in Table 7. Note that, based on the rules in Table 6, 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 a 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. Additionally, 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 7. 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	4.97	2.23	1.69	0.19	3.44	1.46	2.29	2.06	1.53	3.69	0.95	2.13	1.23	0.66
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 it was assigned 1 to be differentiated from T2.

4.1.2. Quality of Phonation

- H1*–H2*;

Figure 5 shows the raw H1*–H2* values of each tone over nine equally 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 (3). Model (3) 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).

$$\text{lmer}(H1^*-H2^* \sim \text{Time} + \text{Tone} + \text{Time} * \text{Tone} + (1 | \text{Participant})) \tag{3}$$

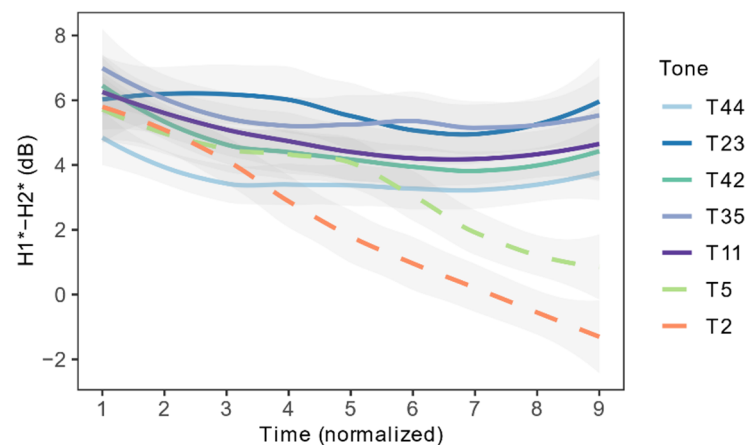


Figure 5. Average H1*–H2* track of the seven tones in Xiapu Min.

The statistics of Model (3) are presented in Tables S8 and S9 in Supplementary Material S3. 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: -0.61 ; T2: -0.92), 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.

- HNR;

Figure 6 shows the raw HNR values of each tone over nine equally timed intervals after averaged across all tokens and all speakers. The HNR contour of checked T2 is below all other tones at every time point except for Point 8. We used linear mixed-effects models to test whether, on average, T2 and T5 have a significantly lower HNR value than unchecked tones (4). Model (4) was run twice, once with T5 and once with T2 as the reference level of Tone. The alpha level was adjusted to 0.025 (0.05/2).

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

The statistics of Model (4) are presented in Tables S10 and S11 in Supplementary Material S3. The results of Model (4) 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 6 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 8 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 and 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 S3. The HNR value predicted by the fitted spline and the first derivative at each time point for each tone are presented in Table S12 in Supplementary Material S3. Table 8 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 the production of T5 targets at a noisy quality toward the end of the vowel.

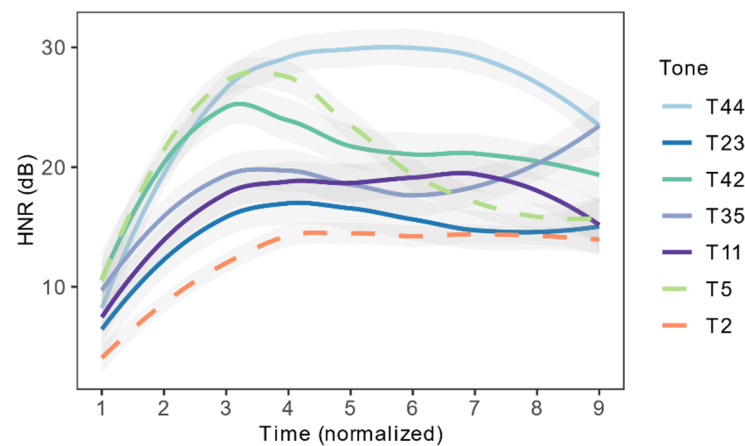


Figure 6. Average HNR track of the seven tones in Xiapu Min.

Table 8. 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	27.45	14.46	28.82	23.32	19.62	16.93	18.82
P9	15.60	13.80	23.26	18.85	23.31	15.29	14.78
P4–P9	11.85	0.66	5.56	4.47	−3.69	1.64	4.04
First derivative at P4	−3.29	0.88	1.59	−2.75	−0.88	0.11	0.07

In summary, checked tones T2 and T5 have more glottal constriction than unchecked tones, as indicated by the 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. Together, the glottal constriction and noisy quality indicate a more glottalized phonatory quality of checked tones compared to unchecked tones. The HNR values do not differentiate checked from unchecked tones as consistently as the H1*–H2* values. Thus, we hypothesize that the listeners are less likely to use aperiodic voicing as a salient cue in differentiating checked tones from unchecked ones in perception. Future studies can manipulate spectral tilt and pulse periodicity separately to test the perceptual saliency of glottal constriction and noise for checked tone identification in Xiapu Min.

In addition to measuring the acoustic parameters, we also inspected the spectrograms of tokens with checked tones to describe the phonatory quality of checked tones compre-

hensively. We find that the realization of the glottal stop in Xiapu Min checked tones varies from strong realization with sustained (voiceless) glottal closure, to moderate realization with voiced glottalization on vowels, to deletion of a glottal stop. Figure 7 shows a spectrogram of tokens of different degree of glottal stop realizations. Out of 236 tokens attested, there are 223 tokens realized with only voiced glottalization or reduced amplitude, 4 with full glottal closure, and 9 with deletion of the glottal stop. We propose that the predominance of glottalized phonation over full glottal closure in checked tones is due to the phrasal position of the target syllables. All the target syllables are elicited in phrase-medial positions. Glottal stops are more likely to be realized as fully voiced in phrase-medial position (Garellek et al. 2021).

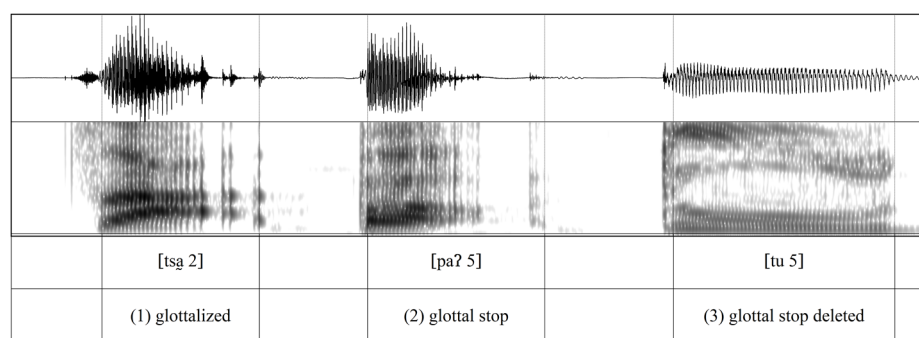


Figure 7. Spectrograms of glottal stops realized as glottalized vowel, full glottal stop and deletion. The underlying form of (1), (2), and (3) is /tsaʔ 2/ 杂 “mixed”, /paʔ 5/ 百 “hundred” and /tuʔ 5/ 督 “supervise”.

4.1.3. Duration

Figure 8 shows the duration of each tone after averaged 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 (5). Model (5) 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 (5) are presented in Tables S13 and S14 in Supplementary Material S3. The results show that both checked T5 and T2 have a significantly shorter duration than every unchecked tone.

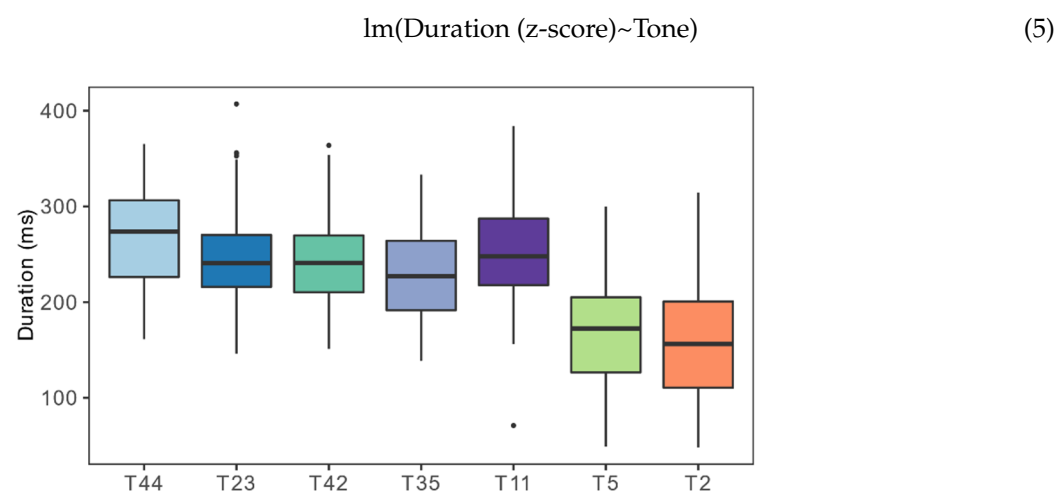


Figure 8. Average duration of the seven tones in Xiapu Min.

4.1.4. 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 diphthongs for comparison to avoid the influence of formant transition in diphthongs. 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 (6) and (7). The statistics of Model (6) and (7) are presented in Tables S15 and S16 in Supplementary Material S3. 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.

$$\text{lmer}(F1 \sim \text{Tone} + \text{Checkedness} + (1 | \text{Participant})) \quad (6)$$

$$\text{lmer}(F2 \sim \text{Tone} + \text{Checkedness} + (1 | \text{Participant})) \quad (7)$$

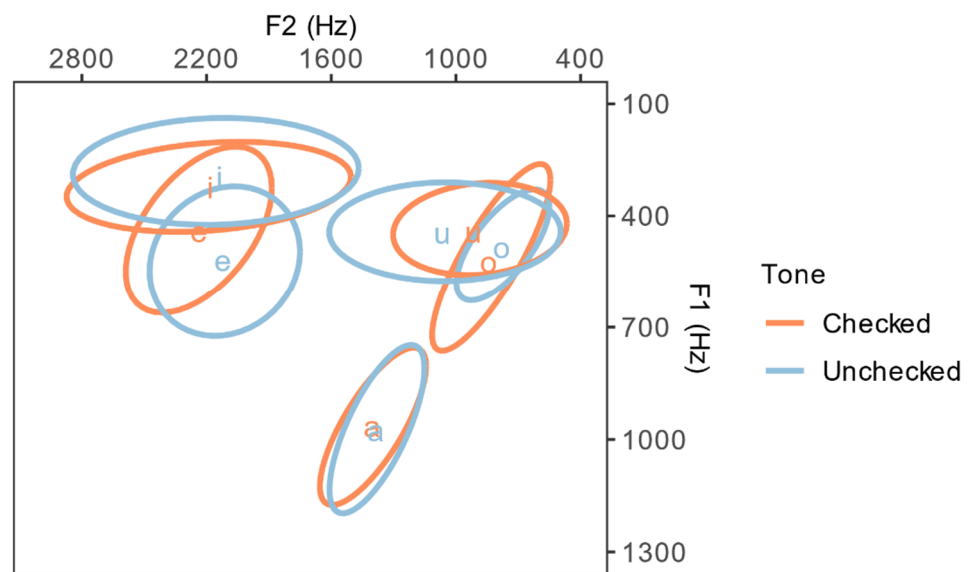


Figure 9. F1 and F2 distribution of vowels in checked and unchecked tones.

In summary, we confirm that T5 and 2 have distinct pitch values from unchecked tones, and we 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 becomes stronger when the production proceeds toward 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 they have distinct F0 values compared to unchecked tones. No significant differences in vowel quality have been found between checked and unchecked tones.

4.2. 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 * two repetitions * ten participants). A total of 66 syllables were excluded because of either corrupted recording or mispronunciation, leaving 754 syllables valid for analyses. VoiceSauce (Shue et al. 2011) yielded 79,100 data points in total. The tracking error and outlier detection and exclusion procedures were the same as described in Section 3.2. After data exclusion, there were 78,960, 74,299, 74,299, 71,638, and 78,617 data points for F0, F1, F2, H1*–H2*, and HNR, respectively. The data points were divided into nine (for plotting the results) and three equally timed intervals (for the linear discriminant analysis). The descriptive statistics of the dataset can be found in Tables S4–S7 in Supplementary Material S3.

4.2.1. Neutralization among T2, T44, and T23

The first sandhi rule of Xiapu Min is {T2, T23, T44} → T44 / ___ X (Table 4, Example g). 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 neutralization. LDA models use a categorical variable as the dependent variable, and they 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 at least one composite linear discriminant score for each token, and it 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 purpose of using LDA models is 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 percentage of the majority class as the chance level, because in random guessing, predicting all the tokens as the majority class results in the highest chance (Bosch and Paquette 2018). The results of the LDA models can help determine whether the neutralization among the three underlyingly different tones is complete or not. The LDA models were implemented by the `lda()` function from the MASS package in R (Venables and Ripley 2002).

The R code for the LDA models is in (8). The dependent variable is the citation tone of the target syllables. The independent variables are the average F0, H1*–H2*, HNR of three equally timed 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.

$$\text{lda}(\text{Tone} \sim \text{F0_1} + \text{F0_2} + \text{F0_3} + \text{H1*} - \text{H2*}_1 + \text{H1*} - \text{H2*}_2 + \text{H1*} - \text{H2*}_3 + \text{HNR_1} + \text{HNR_2} + \text{HNR_3} + \text{Duration}) \quad (8)$$

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. For every LDA model, we calculated its classification accuracy based on a leave-one-out cross-validation.

Figure 10a shows the LD1 and LD2 distribution of T2, T44, and T23 in citation forms. The classification accuracy of the citation forms is 94.81%, which is significantly higher than the 38.96% chance level ($p < 0.001$). We applied the LDA models on each pair of 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 of T23 vs. T44, T2 vs. T23, and T2 vs. T44 in sandhi forms are: 59.46% ($p = 0.31$; chance = 54.05%), 68.69% ($p < 0.001$; chance = 51.52%), and 76.61% ($p < 0.001$, chance = 52.63%). 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 for the other two pairs. Future studies should aim for more balanced stimuli.

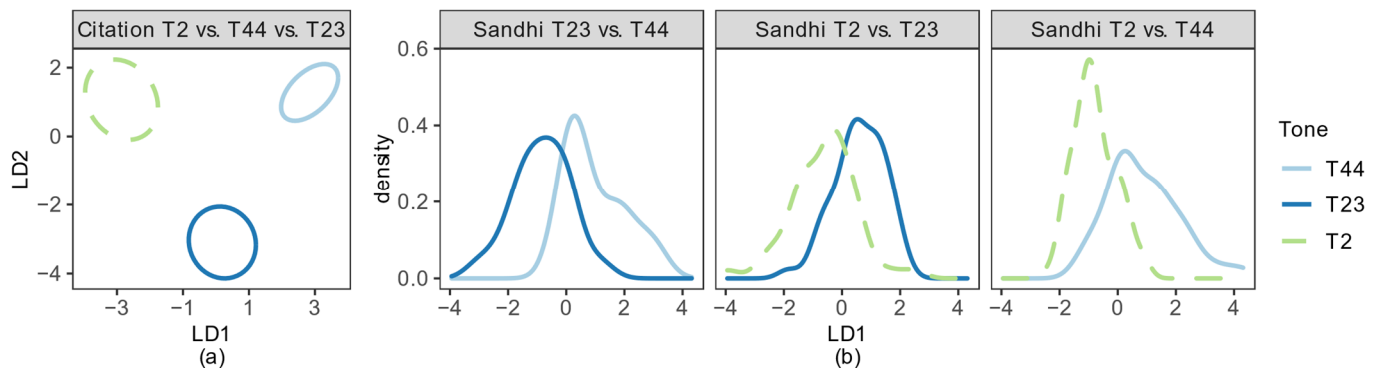


Figure 10. (a) is the first and second linear discriminant score (LD1 and LD2) distribution of T2 vs. T44 vs. T23 in citation forms. The ellipses represent 50% confidence intervals around the mean of each group. (b) is the LD1 distribution of T23 vs. T44, T2 vs. T23, and T2 vs. T44 in sandhi forms, respectively. LD1 and LD2 are from the models without cross-validation.

Next, we ask which acoustic parameters contribute most to the above-chance discriminations. We calculated the Pearson correlation between each acoustic parameter and the linear discriminant scores. For citation tones, LD1 explains 63.3% of the variance. The top three parameters that have the highest absolute correlation with LD1 are **duration**, **mid HNR**, and **final F0**. 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 (8) and the correlations between the parameters and the linear discriminant scores are presented in Tables S17–S21 in Supplementary Material S3.

Figure 11 shows the values of F0, H1*–H2*, HNR, and the 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 that of 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 among those three tones remains, but it 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 S22 and S23 in Supplementary Material S3. In summary, checked T2 has a constricted and noisy quality in citation forms. In sandhi forms, T2 becomes less constricted and less noisy, indicating a reduction of glottalization. The duration of T2 is shorter than that of 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.

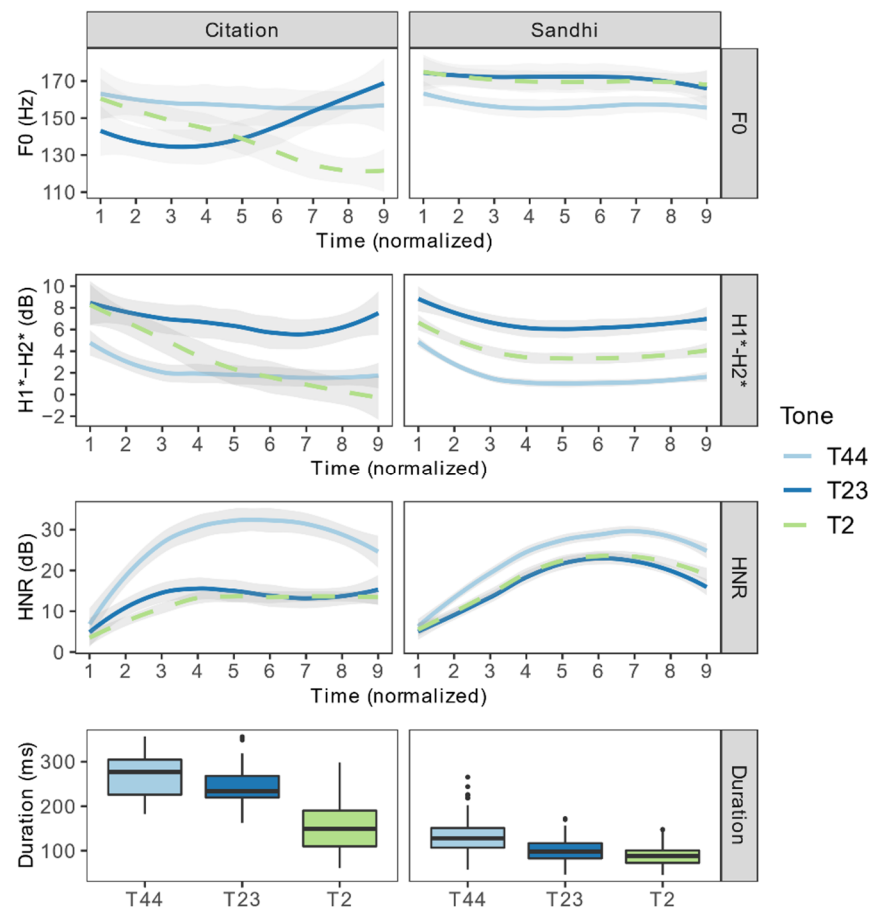


Figure 11. Acoustic parameter values of T44, T23, and T2 in citation and sandhi forms.

4.2.2. Neutralization among T5, T42, and T35

The other sandhi rule of Xiapu Min is {T5, T42, T35} → T55 / ___ X (Table 4, Example h). It results in the neutralization of T5, T42, and T35. Similar to Section 4.2.1, 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 in Formula (8). For every LDA model, we calculated its classification accuracy based on a leave-one-out cross-validation.

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%, which is significantly higher than the 45.57% chance level ($p < 0.001$). 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 of T35 vs. T42, T5 vs. T35, and T5 vs. T42 in sandhi forms are: 47.41% ($p = 0.83$; chance = 51.11%), 80.35% ($p < 0.001$; chance = 51.45%), and 86.79% ($p < 0.001$, chance = 53.77%). The results indicate that, before sandhi, the citation forms of T5, T42, and T35 are differentiated at 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.

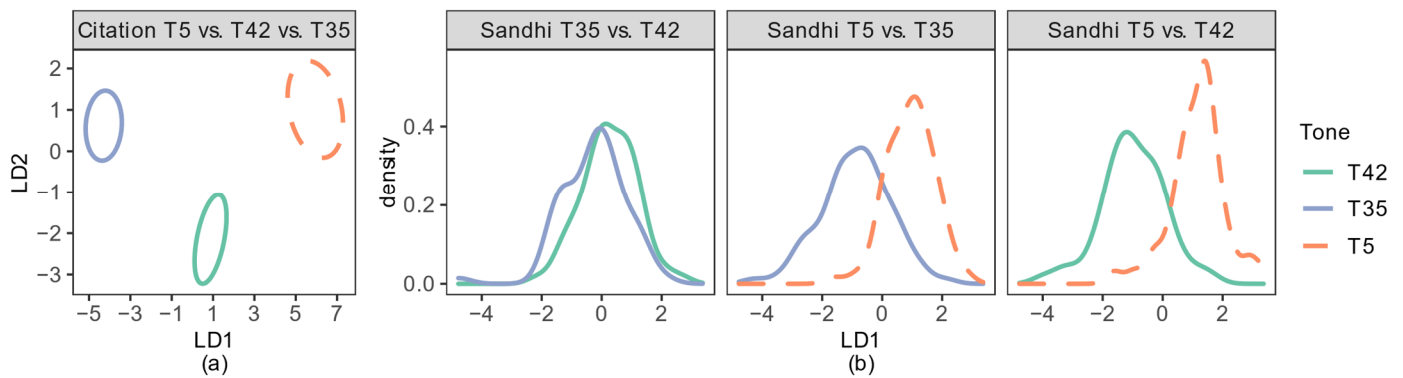


Figure 12. (a) is the LD1-LD2 distribution of T5 vs. T42 vs. T35 in citation forms. The ellipses represent 50% confidence intervals around the mean of each group. (b) is the LD1 distribution of T35 vs. T42, T5 vs. T35, and T5 vs. T42 in sandhi forms, respectively. LD1 and LD2 are from the models without cross-validation.

We calculated the Pearson correlation between each acoustic parameter and the linear discriminant scores to determine which parameters contribute most to the above-chance discriminations. LD1 explains 91.9% 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 S24–S28 in Supplementary Material S3.

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 it 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 it 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 it 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 7. The final 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 presented in Tables S29 and S30 in Supplementary Material S3. 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 reduction 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.

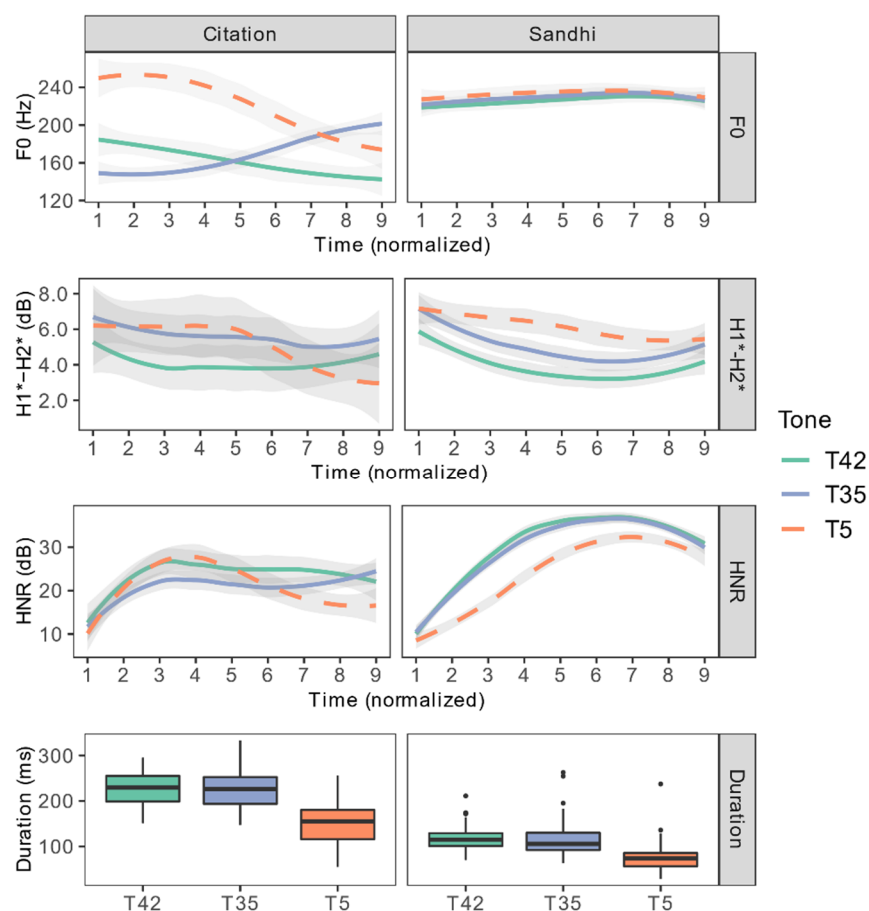


Figure 13. Acoustic parameter values of T42, T35, and T5 in citation and sandhi forms.

4.2.3. Summary of Tonal Neutralization in Sandhi Forms

Table 9 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 four of those 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 acoustic correlate that distinguishes checked tones from unchecked tones in sandhi forms. Table 10 presents the average duration of each tone in citation and sandhi forms. Checked tones remain 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 (citation: 66%; sandhi: 70%).

Among all the sandhi forms, F0 (of the last third of vowels) surfaces as an important correlate only in distinguishing T2 from T23 (Table 9). However, the correlation between F0 and LD score is rather weak (Pearson $r = -0.31$). The absolute difference in final F0 between T23 and 2 is 8 Hz, which is rather small. Thus, we conclude that the F0 difference among checked and unchecked tones in citation forms is largely neutralized in sandhi forms.

Moreover, H1*-H2* (of the first third of vowels) surfaces as an important correlate only when distinguishing T2 from T23 in sandhi forms (Table 9). However, the correlation between H1*-H2* and the LD score is also rather weak (Pearson $r = 0.25$). In citation forms, the H1*-H2* contour of checked tones is in a falling trend (Figures 5, 11 and 13). Checked T2 and 5 end in significantly lower H1*-H2* than unchecked tones. In sandhi forms, however, the H1*-H2* contour of T2 and 5 becomes flatter. The H1*-H2* of T2 is higher than T44. The H1*-H2* of T5 is higher than T35 and 42. Given that a higher H1*-H2* value is correlated with less glottal constriction, we argue that checked tones in

sandhi forms become less constricted, especially at the end of vowels, compared with the citation forms. The difference in glottal constriction between checked and unchecked tones is largely neutralized in sandhi form.

Finally, HNR appears to be an effective parameter when distinguishing T2 from T44, T5 from T35, and T5 from T42 (Table 9). However, for T2 vs. T44, the difference in average HNR becomes smaller in sandhi forms than in citation forms (citation: 15.12 dB; sandhi: 4.88 dB). For T5 vs. T35 and T5 vs. T42, their difference in HNR is in the initial and middle third of the vowel, but not at the end. In contrast, in citation forms, checked T5 is characterized by its steeper falling HNR contour during the latter half of the vowel.

One possible explanation is that HNR differences between T5 and T35 and 42 in the initial two-thirds of the vowel are a by-product of the short duration and the influence of the onset in the checked tones. Three-quarters 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 into the vowels. Additionally, it may also be because vowels bearing checked tones in sandhi forms are extra-short compared to those with unchecked tones (both in citation and sandhi) and to those with checked tones in citation forms. It is possible, then, that checked tones in sandhi forms might 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 suggest that the vowel-final glottalized quality of the checked tones is largely reduced in sandhi forms. In summary, in citation forms, checked tones are differentiated from unchecked tones by having distinct F0 contour, shorter duration, and glottalized quality at the end of the vowels. In sandhi forms, checked tones acquire similar F0 values and phonatory quality to unchecked tones. However, the duration difference between checked and unchecked tones persists.

Table 9. Classification accuracies and top three parameters that have the highest correlation with LD1 (Contrasts that are differentiated significantly above chance are in bold).

Citation/Sandhi	Contrast	Classification Accuracy (Chance Level, <i>p</i> Value)	Parameters Citation/Sandhi
Citation	T2 vs. T44 vs. T23	94.81% (38.96%, <0.001)	duration, mid HNR, final F0
Sandhi	T23 vs. T44	59.46% (54.05%, 0.31)	duration, final F0, initial H1*–H2* duration, initial and final HNR
	T2 vs. T23 T2 vs. T44	68.69% (51.52%, <0.001) 76.61% (52.63%, <0.001)	
Citation	T5 vs. T42 vs. T35	100% (45.57%, <0.001)	initial and mid F0, duration
Sandhi	T35 vs. T42	47.41% (51.11%, 0.83)	duration, initial and mid HNR duration, initial and mid HNR
	T5 vs. T35 T5 vs. T42	80.35% (51.45%, <0.001) 86.79% (53.77%, <0.001)	

Table 10. 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	160	269	243	228	251	251	66%
Sandhi	74	89	133	116	114	100	NA	70%

5. Discussion and Conclusions

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 glottal 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 syllables and vowels that are found in other languages apply to Xiapu Min checked tones. We recommend that future studies on checked syllables and vowels in other languages focus on these four prototypical phonetic properties as well.

We further show how checked tones change when they are phonologically neutralized with unchecked tones. This study finds that incomplete neutralization only occurs 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.

Compared with Taiwanese Min checked tones in sandhi forms (Pan 2017), Xiapu Min behaves in a similar way. Taiwanese Min high and low checked tones [5] and [3] were both produced with more glottal opening than unchecked falling tones in sandhi forms. In Xiapu Min, the spectral tilt of checked tones increases significantly in sandhi forms compared with the citation forms, and it is mostly indistinguishable from that of unchecked tones. Checked tones' sandhi forms are not characterized by a glottalized quality in either Taiwanese or Xiapu Min.

With the acoustic results of checked syllables in citation and sandhi forms, we can also make inferences on the relation between glottalization and short duration in Xiapu Min. In citation forms, we find short duration and glottalization co-occur in CV? syllables. Cross-linguistically, vowels tend to be shorter in closed syllables than in open syllables (Farnetani and Kori 1986; Rietveld and Frauenfelder 1987; van Santen 1992; Maddieson 1985). In general, the realization of glottal stops and voiced glottalization, including in checked syllables, varies across languages (Garellek et al. 2021). Glottal stops are frequently realized as voiced glottalization on adjacent vowels. The co-occurrence of short duration and glottalization in CV? syllables in Xiapu Min thus leads to the following question: are glottalization and short duration due to the coarticulation of closed syllable ending in glottal stop, or do they together comprise a distinct "checked" phonation as opposed to "unchecked" phonation? We argue that neither hypothesis captures the nature of checked syllables in Xiapu Min. First, we argue against the hypothesis that "checked" is a phonation type in Xiapu Min. Table 1 in Section 1 showed that CV? behaves differently from open syllables in phonological transformations, supporting the idea that the glottal stop in CV? syllables in Xiapu Min is a coda segment rather than a suprasegment. Second, we argue that the short duration of checked syllables is independent of their syllable structure. Our evidence comes from the acoustic properties of checked tones in sandhi forms. In sandhi forms, the voice quality of checked tones is similar to that of unchecked tones, indicating that the glottal gesture is reduced or even lost in the surface form. Despite the weakening or loss of glottal gestures, the short duration of checked tones is still preserved in sandhi forms. Based on the above evidence, then, we propose that "checked" is not a phonation type in Xiapu Min. The short duration of checked syllables is not due to the cross-linguistical nature of closed syllables, but it is instead an inherent nature of "being checked" in Xiapu Min.

Based on the acoustic properties of the checked tones in citation and sandhi forms, it is also possible to make predictions about possible sound change patterns of checked syllables in Xiapu Min. Across Chinese languages, there is a tendency for checked syllables to become unchecked (Cao 2002; Feng 2011; Gu 2015; Shen 2007; Song 2009; Xing and Meng 2006; Yang 1982). This phenomenon is referred as "rùshēng shūhuà" (入声舒化 "checked tone opening") in Chinese literature. However, it has been observed that the multiple

features of checked syllables do not disappear at the same time but are lost in sequence. 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 14.

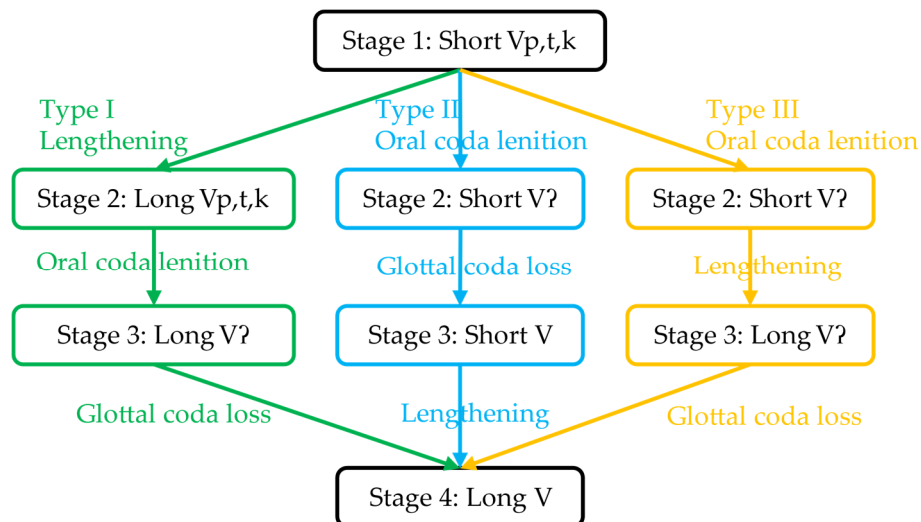


Figure 14. Three paths involving the loss of checkedness (schematized based on Zhu et al. (2008)).

The differences in the three paths depend on the sequencing of vowel lengthening, oral coda lenition, and glottal coda loss. Given how common it is for Chinese languages to lose their checked tones, one might assume that the checked syllables in Xiapu Min are also losing their checkedness. At what stage of the sound change might Xiapu Min currently be? Since Xiapu Min no longer has /p, t, k/ as coda in checked syllables, we can rule out Stage 1 and Stage 2 of Path I. The phonetic features of checked tones in citation forms suggest that Xiapu Min might currently be at Stage 2 of either Path II or Path III because its checked syllables are short and closed by a glottal stop. What might be the next stage of the possible sound change for checked syllables in Xiapu Min? We have observed that glottalization is weakened, whereas the short duration is retained in sandhi forms. This suggests that, in Xiapu Min, duration may be 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 changes in the future, their 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. Such studies will allow us to directly test whether short duration, glottalization, and F0 are independent cues of checked tones in Xiapu Min. 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. Additionally, 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 following supporting information can be downloaded at: <https://doi.org/10.17605/OSF.IO/M5UG2> (accessed on 10 February 2022). Supplementary Material S1: Stimuli list; Supplementary Material S2: Sample recordings of the stimuli; Supplementary Material S3:

Statistical results, supplementary tables and figures; Supplementary Material S4: R code for the statistical analysis; Supplementary Material S5: Data set.

Author Contributions: Conceptualization, Y.C.; Data curation, Y.C. and S.Y.; Formal analysis, Y.C.; Funding acquisition, Y.C.; Investigation, Y.C. and S.Y.; Methodology, Y.C.; Project administration, Y.C.; Resources, Y.C. and S.Y.; Software, Y.C.; Supervision, Y.C.; Validation, S.Y.; Visualization, Y.C.; Writing—original draft, Y.C.; Writing—review & editing, Y.C. and S.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Friends at the International Center at UCSD and the Department of Linguistics at UCSD.

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 California San Diego (protocol code: 190550; date of approval: 29 April 2019).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data are available online at <https://doi.org/10.17605/OSF.IO/M5UG2> (accessed on 10 February 2022) for non-commercial academic or research purposes.

Acknowledgments: We thank Marc Garellek for guiding the study and reviewing the manuscript. We thank Friends of the International Center at UCSD for funding this project. We thank Jinmin Wang for helping us recruit participants. We thank four anonymous reviewers and editors for their constructive feedback and suggestions on the paper. We thank UCSD Phonco and Fieldwork group and the audience at the 178th Meeting of the Acoustical Society of America for their feedback. We thank all the participants for participating in this study.

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

- Ao, Benjamin Xiaoping. 1993. Phonetics and Phonology of Nantong Chinese. Ph.D. dissertation, Ohio State University, Columbus, OH, USA. Available online: https://etd.ohiolink.edu/apexprod/rws_olink/r/1501/10?clear=10&p10_accession_num=osu1487844485897534 (accessed on 15 February 2022).
- Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67: 1–48. [CrossRef]
- Boersma, Paul, and David Weenink. 2021. Praat: Doing Phonetics by Computer [Computer Program] (Version 6.2). Available online: <http://www.praat.org/> (accessed on 14 February 2022).
- Bosch, Nigel, and Luc Paquette. 2018. Metrics for discrete student models: Chance Levels, comparisons, and use cases. *Journal of Learning Analytics* 5: 86–104. [CrossRef]
- Cao, Zhiyun 曹志耘. 2002. Wuhuiyu rusheng yanbian de fangshi 吴徽语入声演变的方式 [On the change of entering tones in Wu and Hui dialects]. *Studies of the Chinese Language* 5: 441–6. Available online: https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2002&filename=YWZG200205006&uniplatform=NZKPT&v=s0B9R-KhbM2fGmUgWpfGAVRwIz2h2Jr5IcZir7Uix_vSZFISSDNnpfFslQ37A_qV (accessed on 14 February 2022).
- Chan, Marjorie K. M. 1987. Tone and melody in Cantonese. *Annual Meeting of the Berkeley Linguistics Society* 13: 26. [CrossRef]
- Chao, Yuen Ren. 1930. ə sistim əv “toun-letəz” [A system of ‘tone-letters’]. *Le Maître Phonétique Troisième Série* 8: 24–27. Available online: <https://www.jstor.org/stable/44704341> (accessed on 14 February 2022).
- 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. Amsterdam: John Benjamins Publishing Company, vol. 2, pp. 147–70. [CrossRef]
- Chien, Yu-Fu, and Allard Jongman. 2019. Tonal neutralization of Taiwanese checked and smooth syllables: An acoustic study. *Language and Speech* 62: 452–74. [CrossRef]
- De Maesschalck, Roy, Delphine Jouan-Rimbaud, and Désiré L. Massart. 2000. The Mahalanobis Distance. *Chemometrics and Intelligent Laboratory Systems* 50: 1–18. [CrossRef]
- Duan, Wenjun, and Yuan Jia. 2015. Contrastive study of focus phonetic realization between Jinan dialect and Taiyuan dialect. Paper presented at the 2015 International Conference Oriental COCOSDA Held Jointly with 2015 Conference on Asian Spoken Language Research and Evaluation (O-COCOSDA/CASLRE), Shanghai, China, October 28–30; pp. 47–52. [CrossRef]
- Esposito, Christina M. 2012. An acoustic and electroglottographic study of White Hmong tone and phonation. *Journal of Phonetics* 40: 466–76. [CrossRef]

- Farnetani, Edda, and Shiro Kori. 1986. Effects of syllable and word structure on segmental durations in spoken Italian. *Speech Communication* 5: 17–34. [CrossRef]
- Feng, Faqiang 冯法强. 2011. Zhongguo rushing zai fangyan zhongde shuhua yanjiu 中古入声在方言中的舒化研究 [Study of the laxation of Middle Chinese Ru tone in modern dialects]. Master's thesis, Shanghai Normal University, Shanghai, China. Available online: <https://kns.cnki.net/kcms/detail/detail.aspx?filename=1011161190.nh&dbcode=CMFD&dbname=CMFD2011&v=OKLZR3sXo2jbu39k1Ggnx6lORkRVju7ifxAKRC9pVhqYACyA7RNF3dgYdHrZPdf> (accessed on 15 February 2022).
- Garellek, Marc. 2019. The phonetics of voice. In *Routledge Handbook of Phonetics*. Edited by William Katz and Peter Assmann. Oxford: Routledge, pp. 75–106.
- Garellek, Marc, and Christina M. Esposito. 2021. Phonetics of White Hmong vowel and tonal contrasts. *Journal of the International Phonetic Association* 6: 1–20. [CrossRef]
- Garellek, Marc, Yuan Chai, Yaqian Huang, and Maxine Van Doren. 2021. Voicing of glottal consonants and non-modal vowels. *Journal of the International Phonetic Association* 7: 1–28. [CrossRef]
- 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: 1078–89. [CrossRef] [PubMed]
- Gruber, James Frederick. 2011. An articulatory, acoustic, and auditory study of Burmese tone. Ph.D. dissertation, Georgetown University, Washington, DC, USA. Available online: <https://repository.library.georgetown.edu/handle/10822/558130> (accessed on 14 February 2022).
- Gu, Qian 顾黔. 2015. Nanjing fangyan Dang Jiang liang sherushengyun de gongshi bianyi ji lishi yanbian yanjiu 南京方言宕江两撮入声韵的共时变异及历时演变研究 [On synchronic variation and diachronic evolution of the entering tone in Dang and Jiang classes in Nanjing dialect]. *Linguistic Sciences* 14: 384–93. Available online: https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2015&filename=YYKE201504006&uniplatform=NZKPT&v=VdwtPmg3bNUB8Cgb2JTkqMMu9PPQ3bisVdBU8dcZLcO5MlJmIabNruVXuKOx_T5 (accessed on 14 February 2022).
- Hu, Na, Yuan Jia, and Bin Liu. 2012. Phonetic and phonological realization of narrow focus in English declarative sentences by Zhenjiang EFL learners. Paper presented at the Speech Prosody 2012, Shanghai, China, May 22–25; pp. 394–97. Available online: https://www.isca-speech.org/archive_v0/sp2012/sp12_394.html (accessed on 14 February 2022).
- Huffman, Marie K. 1987. Measures of phonation type in Hmong. *The Journal of the Acoustical Society of America* 81: 495–504. [CrossRef] [PubMed]
- Izenman, Alan Julian. 2013. Linear discriminant analysis. In *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*. Edited by Alan J. Izenman. Springer Texts in Statistics. New York: Springer, pp. 237–80. [CrossRef]
- Jia, Xiaoying 贾晓莹. 2013. Taiyuan nanjiao xinpai fangyan shengdiao shiyan yanjiu 太原南郊新派方言声调实验研究 [The experimental study on tones in the southern suburb's new-style dialect in Taiyuan]. Master's thesis, Shanxi University, Taiyuan, Shanxi, China. Available online: https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CMFD&dbname=CMFD201401&filename=1013324746.nh&uniplatform=NZKPT&v=k2I_D7XYr0iT6Ytn6U9S2vFmsOwljpv4-K112-8jkFwfQ0yGUw7CioYI52_FutHq (accessed on 14 February 2022).
- Jia, Yuan, and Aijun Li. 2012. Phonetic realization of accent from Chinese English learners in various dialectal regions. Paper presented at the 2012 8th International Symposium on Chinese Spoken Language Processing, Hong Kong, China, December 5–8; pp. 296–300. [CrossRef]
- Keene, Oliver N. 1995. The log transformation is special. *Statistics in Medicine* 14: 811–19. [CrossRef]
- Klatt, Dennis H., and Laura C. Klatt. 1990. Analysis, synthesis, and perception of voice quality variations among female and male Talkers. *The Journal of the Acoustical Society of America* 87: 820–57. [CrossRef]
- 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: EL509–EL515. [CrossRef]
- Kuo, Chen-Hsiu. 2013. Perception and acoustic correlates of the Taiwanese tone sandhi group. Ph.D. dissertation, University of California, Los Angeles, CA, USA. Available online: <https://escholarship.org/uc/item/30q6w11t#main> (accessed on 14 February 2022).
- Lai, Wenpan 赖文盼. 2016. Jiyu Xiamen hua zuo Minnan fangyan shengdiao shiyan yanjiu 基于厦门话作闽南方言声调实验研究 [Acoustic analysis of tones in Southern Min dialect based on Xiamen dialect]. *Art Science and Technology* 29: 129–30. Available online: https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2016&filename=YSKK201604118&uniplatform=NZKPT&v=SzxIjrhYH9kwNEydmNu85tiU6iQChwxj03GkVDzAIs6e4waB_xcHYsgJX_r3jLhB (accessed on 14 February 2022).
- Lalhmingshui, Wendy, and Priyankoo Sarmah. 2018. Production and perception of rising tone sandhi in Mizo. Paper presented at the 6th International Symposium on Tonal Aspects of Languages (TAL 2018), Berlin, Germany, June 20; pp. 114–18. [CrossRef]
- Li, Bing 李兵, and Yanni Liu 刘彦妮. 2006. Changsha fangyan danzidiao ji biandiao de shiyan yuyinxue baogao 长沙方言单字调及变调的实验语音学报告 [An acoustic study of citation tones and tone sandhi in Changsha Chinese]. *Journal of Hunan University (Social Sciences)* 20: 107–12. Available online: https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2006&filename=HDXB200604019&uniplatform=NZKPT&v=2y1mAjUFLfJk2_o-UXRuG9ncebxJ-LIikFXmrGb9H4dKxve88sV_3p3oyIK8dYe (accessed on 14 February 2022).
- Li, Yang. 2016. Complete and incomplete neutralisation in Fuzhou tone sandhi. Paper presented at the 5th International Symposium on Tonal Aspects of Languages (TAL 2016), Buffalo, NY, USA, May 24–27; pp. 116–20. [CrossRef]

- Lin, Hwei-Bing, and Bruno H. Repp. 1989. Cues to the perception of Taiwanese tones. *Language and Speech* 32: 25–44. [CrossRef]
- Liu, Lili 刘俐李. 2008. Jipin guiyi he diaoxi guizheng de fangyan shiyan 基频归一和调系归整的方言实验 [F0 normalization and tone adjustment in dialect experiments]. *Chinese Journal of Phonetics* 1: 221–27. Available online: <https://d.wanfangdata.com.cn/conference/7531003> (accessed on 14 February 2022).
- Liu, Zhangcai 刘掌才. 2013. Hunan Yuanjiang Chishan hua danzidiao he shuangzidiao shengxue shiyan yanjiu 湖南沅江赤山话单字调和双字调声学实验研究 [An acoustic study of mono-syllable tone and tone sandhi of Yuanjiang Chishan dialect in Hu-Nan Province]. Master's thesis, Guangxi Normal University, Guilin, Guangxi, China. Available online: <https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CMFD&dbname=CMFD201402&filename=1013245249.nh&uniplatform=NZKPT&v=5HIyN8FSwIE3J8WTNeVYPZZ6IIPpXy1SfH1OvY8GeADP9TrbCL5u2Ga5Pi%25mmd2FiNpmS> (accessed on 14 February 2022).
- Maddieson, Ian. 1985. Phonetic cues to syllabification. In *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*. Edited by Victoria Fromkin. Orlando: Academic Press, pp. 203–21.
- Michaud, Alexis. 2004. Final consonants and glottalization: New perspectives from Hanoi Vietnamese. *Phonetica* 61: 119–46. [CrossRef] [PubMed]
- Oakden, Christopher. 2017. Checked tone merger in the Nanjing dialect: An acoustic analysis. Paper presented at the 29th North American Conference on Chinese Linguistics (NACCL-29), New Brunswick, NJ, USA, June 16–18; Edited by Lan Zhang. Columbus: Ohio State University, vol. 1, pp. 141–52. Available online: <https://naccl.osu.edu/sites/naccl.osu.edu/files/06-Oakden-p141-152.pdf> (accessed on 14 February 2022).
- Pan, Ho-Hsien. 2005. Voice quality of falling tones in Taiwan Min. Paper presented at the INTERSPEECH 2005, Lisbon, Portugal, September 4–8; pp. 1401–4. [CrossRef]
- Pan, Ho-Hsien. 2017. Glottalization of Taiwan Min checked tones. *Journal of the International Phonetic Association* 47: 37–63. [CrossRef]
- Pan, Ho-Hsien, Hsiao-Tung Huang, and Shao-Ren Lyu. 2016. Coda stop and Taiwan Min checked tone sound changes. Paper presented at the Interspeech 2016, San Francisco, CA, USA, September 8–12; pp. 1011–15. [CrossRef]
- Pan, Ho-Hsien, and Shao-Ren Lyu. 2021. Taiwan Min Nan (Taiwanese) checked tones sound change. Paper presented at the Interspeech 2021, Incheon, Korea, September 4–10; pp. 2641–45. [CrossRef]
- 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: 195–203. [CrossRef]
- Pickett, Velma B., María Villalobos Villalobos, and Stephen A. Marlett. 2010. Isthmus (Juchitán) Zapotec. *Journal of the International Phonetic Association* 40: 365–72. [CrossRef]
- 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* 3. [CrossRef]
- Ramsey, Jim, and Brian Ripley. 2017. Pspine: Penalized Smoothing Splines (version 1.0-18). Available online: <https://CRAN.R-project.org/package=pspline> (accessed on 14 February 2022).
- 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.
- Rietveld, A. C. M., and Ulrich Hans Frauenfelder. 1987. The effect of syllable structure on vowel duration. Paper presented at the 11th International Congress of Phonetic Sciences, Tallinn, Estonia, August 1–7; pp. 28–31. Available online: <https://archive-ouverte.unige.ch/unige:85564> (accessed on 14 February 2022).
- Seyfarth, Scott, and Marc Garellek. 2018. Plosive voicing acoustics and voice quality in Yerevan Armenian. *Journal of Phonetics* 71: 425–50. [CrossRef]
- Shao, Dandan 邵丹丹. 2012. Jiyu EGG de Meixian, Fuzhou, Changsha fangyan shengdiao shiyan yanjiu 基于EGG的梅县、福州、长沙方言声调实验研究 [Acoustic experimental analysis of Meixian, Fuzhou, and Changsha dialect based on EGG data]. Master's thesis, Nanjing Normal University, Nanjing, Jiangsu, China. Available online: <https://cdmd.cnki.com.cn/Article/CDMD-10319-1013105547.htm> (accessed on 14 February 2022).
- Shen, Ming 沈明. 2007. Jinyu Wutaipian rushengdiaode yanbian 晋语五台片入声调的演变 [On the change of entering tone of Wutai cluster in Jin group]. *Dialect* 4: 296–304. Available online: <https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2007&filename=FYZA200704005&uniplatform=NZKPT&v=wf-ZZcpwLvonVi4nagZA0mZiAFcKj3v8oIFGbbKdwDY7714Rn9aAKmJL9UoM3kYH> (accessed on 14 February 2022).
- Shi, Feng 石锋, Qibin Ran 冉启斌, and Ping Wang 王萍. 2010. Lun yuyin geju 论语音格局 [On sound pattern]. *Nankai Linguistics* 1: 1–14. Available online: <https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2010&filename=NKYK201001003&uniplatform=NZKPT&v=p0Tr3X2zKSpMI5HT22HT-fCWqMvnnNtHfVGTpv5Y7AhQGdhva8eVmlKJvTLxwn> (accessed on 14 February 2022).
- Shue, Yen-Liang, Patricia A. Keating, Chad Vicenik, and Kristine Yu. 2011. VoiceSauce: A program for voice analysis. Paper presented at the International Congress of Phonetic Sciences, Hong Kong, China, August 17–21; pp. 1846–9.
- Song, Yidan 宋益丹. 2009. Nanjing fangyan zhongde rusheng housewei shiyan yanjiu 南京方言中的入声喉塞尾实验研究 [On experiment of entering tone of Nanjing dialect]. *Journal of School of Chinese Language and Culture Nanjing Normal University* 2: 166–72. Available online: <https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2009&filename=NJSD200902038&uniplatform=NZKPT&v=15kAM0Iw8dEgT8wCtY5IrXI-R3W-uwijq8pgFT0OpGurM5sp6yW-YxmRmJgMRbXu> (accessed on 14 February 2022).

- Speck, Charles H. 1978. The phonology of Texmelucan Zapotec verb irregularity. Master's thesis, University of North Dakota, Grand Forks, ND, USA. Available online: <https://commons.und.edu/theses/2660> (accessed on 14 February 2022).
- Su, Hui 苏慧. 2016. Jiangdu fangyan danzidiao shiyan yanjiu 江都方言单字调实验研究 [Experimental analysis of Jiangdu Dialect citation tone]. *Modern Chinese* 4: 19–23. Available online: <https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLASN2019&filename=YWLY201604007&uniplatform=NZKPT&v=VSM7oxLJkqm9-zTTT9doAtYGGQQ5ww3qxS-gq0D-5Q9Vda52r5ALVK59K8B-B8Up> (accessed on 14 February 2022).
- Sun, Huaxian 孙华先. 2003. Nanjing fangyan shengdiao de ruogan wenti 南京方言声调的若干问题 [A discussion on the tones of Nanjing dialect]. *Journal of Nanjing Xiaozhuang College* 19: 34–40. Available online: <https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2003&filename=SFZK200301007&uniplatform=NZKPT&v=waY0hHh1so6jOtKWtQwOxoUZVxal1L5madqKZRzULYFF3ezEOJK0XyhT350WVG8z> (accessed on 14 February 2022).
- Tang, Zhiqiang 唐志强. 2014. Wanshu Jianghuai guanhua rusheng shiyan yanjiu 皖属江淮官话入声实验研究 [Experimental analysis of checked tone in Anhui Jianghuai Mandarin]. Master's thesis, Nanjing Normal University, Nanjing, Jiangsu, China. Available online: <https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CMFD&dbname=CMFD201501&filename=1014347991.nh&uniplatform=NZKPT&v=iPwMUEH7qe0ikdrD9kZAlc%25mmd2F2xJSf61Ck42RSLQm0obCKMmGjvuBULmEjVjJHQb1> (accessed on 14 February 2022).
- Teodocio Olivares, Amador. 2009. Betaza Zapotec phonology: Segmental and suprasegmental features. Master's thesis, University of Texas at Austin, Austin, TX, USA. Available online: <https://repositories.lib.utexas.edu/handle/2152/19162> (accessed on 14 February 2022).
- Venables, William N., and Brian D. Ripley. 2002. *Modern Applied Statistics with S*, 4th ed. Statistics and Computing. New York: Springer.
- van Santen, Jan P. H. 1992. Contextual effects on vowel duration. *Speech Communication* 11: 513–46. [CrossRef]
- Wen, Jing 温静. 2015. Mindongqu Xiapu fangyan yuyin yanjiu 闽东区霞浦方言语音研究 [A study on the phonology of Xiapu dialect]. Master's thesis, Fujian Normal University, Fuzhou, Fujian, China. Available online: <https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CMFD&dbname=CMFD201601&filename=1015720187.nh&uniplatform=NZKPT&v=taJQvNmlO3IpPKCgVcRTtW2y8H1jQjifrIuyGsKi9UELK4n6ZFJ2i5dENKvLXYqr> (accessed on 14 February 2022).
- Wu, Bo. 2018. An Acoustic Analysis of Vowels in Checked Syllables in Chinese. *Chinese Journal of Acoustics* 4: 491–502. [CrossRef]
- Xia, Liping, and Fang Hu. 2016. Vowels and diphthongs in the Taiyuan Jin Chinese dialect. Paper presented at the Interspeech 2016, San Francisco, CA, USA, September 8–12; pp. 993–7. [CrossRef]
- Xiapu County Bureau of Statistics 霞浦县统计局. 2021. Xiapu Xian Diqici Quanguo Renkou Pucha Gongbao 霞浦县第七次全国人口普查公报 [Xiapu County Seventh National Census Report]. Available online: http://www.xiapu.gov.cn/zwgk/zfxgkzdgz/tjxx/tjgb/202106/t20210628_1491098.htm (accessed on 14 February 2022).
- Xing, Xiangdong 邢向东, and Wanchun Meng 孟万春. 2006. Shaanbei Ganquan Yanchang fangyan rushengzi duyin yanjiu 陕北甘泉、延长方言入声字读音研究 [On the pronunciation of entering tone characters in Ganquan and Yanchang dialects in Northern Shaanxi]. *Chinese Language* 5: 444–51. Available online: https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2006&filename=YWZG200605011&uniplatform=NZKPT&v=J4AHOlBFZukas_X-6IooOyGPOkhi-_pqKiu4dkxejIq0YINxBXeitln934OAjdv (accessed on 14 February 2022).
- Yang, Shuzu 杨述祖. 1982. Shanxi fangyan rushengde xianzhuang jiqi fazhan qushi 山西方言入声的现状及其发展趋势 [Synchronic distribution of entering tone in Shanxi dialects and their evolution tendency]. *Linguistic Research* 1: 130–3. Available online: https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD7984&filename=YWYJ198201021&uniplatform=NZKPT&v=GCPiTBeyAbjxyVRY7Uyiud0ISxsB_ISQ_gwQLjkr2bka1x4-aQbvam_UqGiTq8pj (accessed on 14 February 2022).
- 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. Lecture Notes in Computer Science. Cham: Springer International Publishing, vol. 10733, pp. 146–57. [CrossRef]
- Zee, Eric, and Ian Maddieson. 1979. Tones and tone sandhi in Shanghai: Phonetic evidence and phonological analysis. *UCLA Working Papers in Phonetics* 45: 93–129.
- Zhang, Zhaoyan. 2016. Cause-effect relationship between vocal fold physiology and voice production in a three-dimensional phonation model. *The Journal of the Acoustical Society of America* 139: 1493–1507. [CrossRef]
- Zhang, Jingfen. 2020. The evolution of checked tones. In *Tono-Types and Tone Evolution*. Frontiers in Chinese Linguistics. Edited by Jingfen Zhang. Singapore: Springer, vol. 11, pp. 121–50. [CrossRef]
- Zhu, Xiaonong 朱晓农. 2004. Jipin guiyihua—Ruhe chuli shengdiao de sui ji chanyi? 基频归一化—如何处理声调的随机差异? [F0 normalization: How to deal with between speaker tonal variations?]. *Linguistic Sciences* 3: 3–19. Available online: http://journal15.magtechjournal.com/Jwk_yyxx/CN/Y2004/V3/I2/3 (accessed on 14 February 2022).
- Zhu, Xiaonong 朱晓农, Lei Jiao 焦磊, Zhicheng Yan 严至诚, and Ying Hong 洪英. 2008. Rusheng yanhua santu 入声演化三途 [Three ways of Rusheng sound change]. *Studies of the Chinese Language* 4: 324–38. Available online: https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2008&filename=YWZG200804007&uniplatform=NZKPT&v=Bu-eB5OmdI4kcpQO6y-Sw6wtkb06DDs_OhU-7scdVxkwr8TYeg9JgiODsyRRBwM (accessed on 14 February 2022).