Title
Phonetics and phonology of checked phonation, syllables, and tones

Author
Chai, Yuan

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Phonetics and phonology of checked phonation, syllables, and tones

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Linguistics

by

Yuan Chai

Committee in charge:

Professor Marc Garellek, Chair
Professor Gabriela Caballero
Professor Sarah Creel
Professor Sharon Rose
Professor Will Styler

2022

The Dissertation of Yuan Chai is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Marc Garellek. Marc let me know what kind of scholar I want to be, and what kind of advisor I hope to become. When I experience failures and difficulties, he encourages me to learn from them and helps me overcome them. For every small step that I moved forward, he was there celebrating it. None of my achievements during my Ph.D. study could have been possible without Marc's guidance.

I would like to thank my other committee members: Dr. Sarah Creel, Dr. Gabriela Caballero, Dr. Sharon Rose, and Dr. Will Styler. I am grateful for their guidance and advice throughout the process of formulating this dissertation, and during my graduate study.

The field research in this dissertation was supported by the research grants from Friends of the International Center and the Department of Linguistics at UCSD.

I would like to attribute my gratitude to Ye Shihong, for being my college roommate, research collaborator, colleague in teaching, and friend. Shihong inspired me to study her native language – Xiapu Min, and provided me with every resource she can find. Without her, half of this dissertation wouldn't have been possible. I would also like to thank her family for being my language consultants and helping me recruit participants.

I thank my language teachers of Yateé Zapotec, Adrian Fernández and Briseida Mendez. They have been my consultants of Yateé Zapotec for two years, and have involved me in the community's project of documenting the language. Our collaboration deepened my understanding of Zapotec in general, and is greatly beneficial for my survey of Zapotec languages in this dissertation.

My thanks also goes to my research assistants: Grace Kim, Kaylina Moreno, Marilyn Sun, Chunxiao Zhang, and Zhien Zhang. They helped me collect, synthesize, and analyze the data for the surveys of Zapotec and Chinese languages. I appreciate the hard work and original ideas that they have contributed to the studies. I thank all my participants for participating in the experiments.

Duk-Ho Jung and Nina Hagen Kaldhol, thank you for supporting me during my difficult times, and lifting me up from the low points of my life. Thanks to Yaqian Huang, for spending time with me during the past five years. Thanks to Maxine Van Doren, for being a great lab mate. Thanks to Vivian Liu, for helping me and making time for me whenever I asked. Thanks to Zhang Yiyi, for helping me solve technical problems. Thanks to Yang Zhonglin, for her care and kindness. Thanks to Zhao Yiding, for having been my friend for eighteen years, and continuing being with me on this road. Her unconditional support motivates me to keep going.

I thank my parents, Chai Shufan and Jiang Hong, for supporting me throughout my graduate studies, and respecting my choices and beliefs. I want to thank my grandparents, Lin Xiujie and Jiang Benxiang, who are the first teachers in my life. May my grandmother Lin Xiujie rest in peace.

Chapters 4 and 5, in part, are reprints of the material as it appears in Chai, Yuan, and Shihong Ye. 2022. "Checked Syllables, Checked Tones, and Tone Sandhi in Xiapu Min" *Languages* 7, no. 1: 47. The dissertation author was the primary investigator and author of this paper.

Chapters 6 and 7, in full, are currently being prepared for submission for publication of the material. Chai, Yuan, and Shihong Ye. "Acoustic cues for checked tone perception in Xiapu Min". The dissertation author was the primary investigator and author of this material.

VITA

2015	Bachelor of Arts, Beijing Normal University
2017	Master of Arts, University of Colorado Boulder
2017–2022	Teaching assistant, University of California San Diego

PUBLICATIONS

Chai, Yuan, and Shihong Ye. 2022. "Checked Syllables, Checked Tones, and Tone Sandhi in Xiapu Min." *Languages* 7 (1): 47.

Chai, Yuan, Titus Kubri Kajo Kunda, Alejandro Rodríguez, and Sharon Rose. 2022. "The prosody of declaratives and questions in Rere (Koalib)." In *Prosody and Prosodic Interfaces*, edited by Haruo Kubozono, Junko Ito, and Armin Mester, 298–336. Oxford, UK: Oxford University Press.

Caballero, Gabriela, Yuan Chai, and Marc Garellek. 2022. "Stress, tone, and intonation in Choguita Rarámuri." In *Prosody and Prosodic Interfaces*, edited by Haruo Kubozono, Junko Ito, and Armin Mester, 227–248. Oxford, UK: Oxford University Press.

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*, 1–28.

Chai, Yuan. 2019. "The source of creak in Mandarin utterances." In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, edited by Sasha Calhoun, Paola Escudero, Marija Tabain, and Paul Warren, 1858–1862. Canberra, Australia: Australasian Speech Science / Technology Association Inc.

ABSTRACT OF THE DISSERTATION

Phonetics and phonology of checked phonation, syllables, and tones

by

Yuan Chai

Doctor of Philosophy in Linguistics

University of California San Diego, 2022

Professor Marc Garellek, Chair

The term "checked" has been used to describe constituents whose voicing ends abruptly, often with glottalization. However, there is still no phonological definition of "checked." This dissertation aims to define what "checked" means phonologically, and to describe the phonetic nature of checked constituents in production and perception.

The first part of the dissertation focuses on what it means to be "checked" in two language families where the term is frequently invoked: Zapotec and Chinese. Surveying checked vowels in 20 Zapotec languages, I argue that, in Zapotec, "checked" is a phonation type, represented as a late-phased [+constricted glottis] feature on vowels. Checked phonation notably contrasts with rearticulated phonation with earlier glottalization. Moreover, checked phonation in most

Zapotec languages contrasts with unchecked phonations independently of tone.

Surveying checked syllables in 95 Chinese languages, I argue that, in Chinese, "checked" is a descriptor of the phonotactic constraint between certain closed syllables and tones. Checked syllables are those closed by obstruent codas and bearing different tones from open and sonorant-closed syllables. The tones borne by checked syllables are called "checked tones." The checked syllables/tones in Modern Chinese have developed from the **Rù* syllables/tones in Middle Chinese.

The second part of the dissertation explores the acoustics of checked syllables and tones in Xiapu Min, a Chinese language. I find that, in the citation forms, checked syllables/tones in Xiapu Min have different F0, shorter duration, and more glottalized quality compared with unchecked tones. In sandhi forms where checked tones neutralize with unchecked tones, checked tones nonetheless remain shorter than unchecked tones.

In a series of experiments, I describe the cues the listeners use to identify checked syllables and tones in Xiapu Min. I show that F0, duration, and voice quality each has an independent effect on eliciting checked tone responses, and that listeners can distinguish checked tones from unchecked tones in sandhi forms by the shorter duration of checked tones.

This dissertation provides criteria for future research to determine whether a language has checked constituents. Also, it contributes novel data from an under-documented language to the phonetics of checked syllables and tones.

Chapter 1

Introduction

Generally, a phonological unit described as "checked" has certain phonetic characteristics: short duration, final glottalization or final stop closure, and sometimes specific F0 contours. However, these phonetic characteristics are each already associated with other phonological features/categories: short duration (phonological length), glottalized phonation, closed syllables, and tones. Therefore, the central questions of this dissertation are: 1) at the phonological level, (when) is it necessary to posit a phonological constituent called "checked"? In other words, when does "checked" refer to a distinct phonological constituent that cannot be covered by any of these existing phonological constituents? 2) for the "checked" constituent that refers to a distinct phonological constituent, what are its phonetic properties, and what acoustic cues do listeners use to perceive it? The first question will be answered based on the results of surveys of Zapotec and Chinese languages. I answer the second question through a case study of an under-documented Chinese language – Xiapu Min – using my own fieldwork data.

Before delving into the survey of languages that have a checked constituent, I first review four example languages from different families or sub-families – Isthmus Zapotec (Otomanguean), Meixian Hakka (Sino-Tibetan, Chinese), Vietnamese (Austroasiatic), and Burmese (Sino-Tibetan, Lolo-Burmese). Because there is debate over the phonological structure that the checked feature docks onto, I will refer to the constituent that bears the checked feature as a "checked constituent" when discussing the checked feature in different languages in general. These languages have each been described as having a "checked" constituent. I put quotation marks around the word "checked" when quoting the "checked" contituents identified in existing studies. By examining the phonological and phonetic properties of the "checked" constituents in those languages, we can see what it means to be called "checked" in the literature, what similarities and differences exist across so-called "checked" constituents across languages, and whether a unified cross-linguistic definition can be formulated to describe checked constituents. All examples from these five languages are drawn from the recordings in *Journal of International Phonetic Association* illustrations. The recordings can be found at https://www.internationalphoneticassociation.org/member/audio-files-illustrations-ipa.

1.1 "Checked vowels" in Isthmus Zapotec

Pickett, Villalobos, and Marlett (2010) report that Isthmus Zapotec has three phonations: modal, laryngealized, and "checked." Vowels with "checked" phonation are called "checked" vowels. As described in Pickett, Villalobos, and Marlett (2010, p. 367), "checked" vowels "typically sound like a vowel that ends in a glottal stop", and "may also be slightly laryngealized at times preceding the abrupt glottal closure." In contrast, laryngealized vowels are produced either with creaky voice throughout the vowel, or with a weak glottal closure in the middle of the vowel. A near-minimal pair of checked, laryngealized, and modal phonations is in Example (1). The spectrograms of the three words in the near-minimal pair are in Figure 1.1.

(1)	Checked	/gì^{ʔ}/[a]	"excrement"
	Laryngealized	/ʒḭ̀/	"nose"
	Modal	/gì/	"fire"

(Pickett, Villalobos, and Marlett 2010, p. 368; V̀ represents a low-toned vowel.)

[a] The word is transcribed as [gìʔ] in Pickett, Villalobos, and Marlett 2010. I transcribe it as a superscript because Pickett, Villalobos, and Marlett 2010 maintains that Vʔ is a phonation.

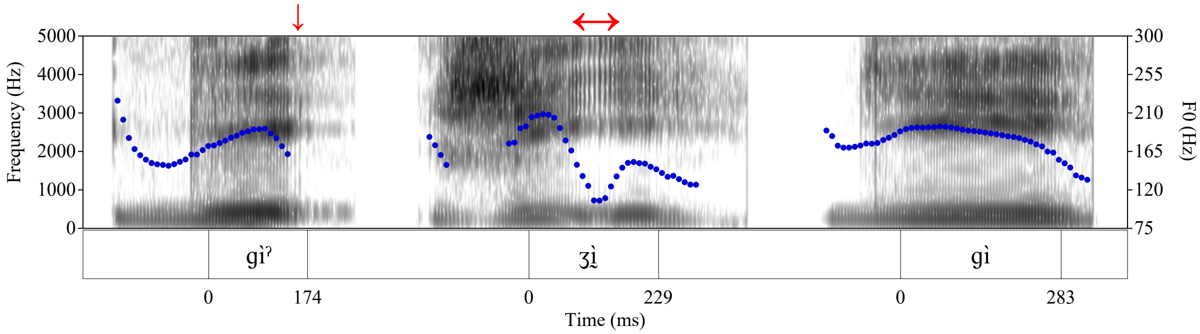


Figure 1.1. Spectrograms of checked, laryngealized, and modal vowel in Isthmus Zapotec. The red arrows mark the glottalized portions. The blue lines represent the F0.

As shown in Figure 1.1, the end of the "checked" vowel has reduced amplitude and widely and irregularly-spaced pulses. The laryngealized vowel has low amplitude and low F0 in the middle of the vowel. In the modal vowel, the pulses are evenly-spaced, and there is no abrupt fall in the amplitude. The spectrograms of tokens with the three contrastive phonation types in Isthmus Zapotec fit their descriptions in Pickett, Villalobos, and Marlett 2010: "checked" vowels in Isthmus Zapotec have final-glottalization, whereas larygnealized vowels have middle-glottalization. Among the three examples, their durations rank as checked < laryngealized < modal.

1.2 "Checked syllables" in Meixian Hakka

Lee and Zee (2009) report that Meixian Hakka has "checked" syllables. "Checked" syllables are syllables that are closed by /-p, -t, -k/. They can only bear high-level (55) or mid-falling (31) tones[1], whereas open syllables and nasal-closed syllables can bear four tones: mid-level (33), high-falling (53), mid-falling (31), or low-level (11) tones. Minimal pairs of the six tones are shown in Example (2). The spectrograms of the two checked tones and the unchecked mid-falling tone (31) are presented in Figure 1.2.

1. The tones values are in Chao numerals (Chao 1930). The first numeral indicates the onset pitch level of the tone. The second numeral indicates the offset pitch level of the tone.

(2) "Checked" high-level it 55 "wing"

"Checked" high-level	it 55	"wing"
"Checked" mid-falling	it 31	"benefit"
mid-falling	i 31	"rain"
mid-level	i 33	"clothes"
high-falling	i 53	"idea"
low-level	i 11	"aunt"

(Lee and Zee 2009, p. 109)

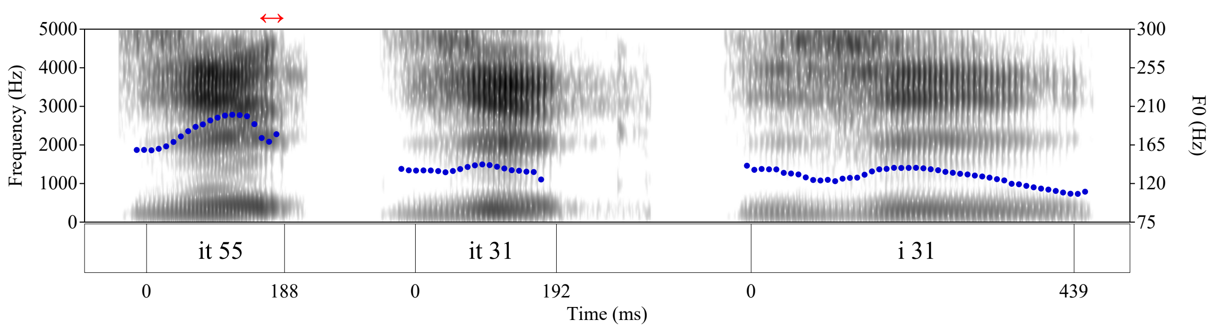


Figure 1.2. Spectrogram of checked T55, 31, and unchecked T31 in Hakka. The red arrows mark the glottalized portions. The blue lines represent the F0.

As the spectrograms in Figure 1.2 show, the "checked" high-level T55 (/it 55/) does not have a visible release of [t] after the vowel [i]. The end of the vowel [i] is glottalized, as indicated by the widely-spaced pulses in the spectrogram. The "checked" mid-falling T31 (/it 31/) has released /t/ coda. There is no visible glottalization in the vowel. Unchecked T31 (/i 31/) has a modal voice quality in the vowel, except that there is a frication noise at the beginning of the vowel[2]. The "checked" T55 and T31 both have shorter duration than the "unchecked" T31.

Comparing Meixian Hakka "checked" syllables with Isthmus "checked" phonation, we see that phonetically, the "checked" constituents in both languages are shorter than the "unchecked" constituents in terms of their duration. In terms of voice quality, Isthmus "checked" vowel ends with glottalization, whereas Mexian Hakka can have a modal voice quality before

2. It is unclear whether the frication noise is consistently produced in the "unchecked" tones, or in the vowel /i/.

the closure of the oral stop coda. In terms of the phonological structure, "checked" in Isthmus Zapotec refers to a phonation type, whereas in Mexian Hakka, "checked" is a type of syllable and the tones that are associated with that syllable.

1.3 "Checked tones" in Vietnamese

Similiar to Meixian Hakka, Vietnamese (Hanoi dialect of Northern Vietnamese) is also reported to have "checked" syllables: /Vp, -t, -k/; and two "checked" tones: D1 sắc (45) and D2 nặng (21), which are borne by "checked syllables" (Kirby 2011; Michaud 2004). The difference between Meixian Hakka and Hanoi Vietnamese is that certain Vietnamese "unchecked" tones are glottalized. Among the six "unchecked" tones in Vietnamese, B2 nặng (22) tone ends with strong glottalization, while C2 ngã (325) has strong glottalization in the middle. All the "unchecked" tones are borne by open or nasal-closed syllables. A minimal pair of the eight tones in Vietnamese is in (3). The spectrograms of words with the two "checked" tones D1 and D2 and the glottalized "unchecked" tones B2 and C2 in (3) are presented in Figure 1.3.

(3)	D1 sắc (45)	"Checked" rising	/mat 45/	"cool"
	D2 nặng (21)	"Checked" low	/mat 21/	"louse, bug"
	A1 ngang (44)	level	/ma 44/	"ghost"
	A2 huyền (32)	mid-falling 31	/ma 32/	"but, yet"
	B1 sắc (24)	rising	/ma 24/	"cheek"
	B2 nặng (22)	low glottalized	/maˀ 22/[a]	"rice seedling"
	C1 hỏi (312)	low-falling	/ma 312/	"tomb"
	C2 ngã (325)	broken	/maˀa 325/[b]	"code"

(Kirby 2011, p. 386)

[a]Kirby (2011) transcribed this word as /ma 22/. I added the glottalization to the endof the vowel to indicate its phonation type.

[b]Kirby (2011) transcribed this word as /ma 325/. I added the glottalization to the middle of the vowel to indicate its phonation type.

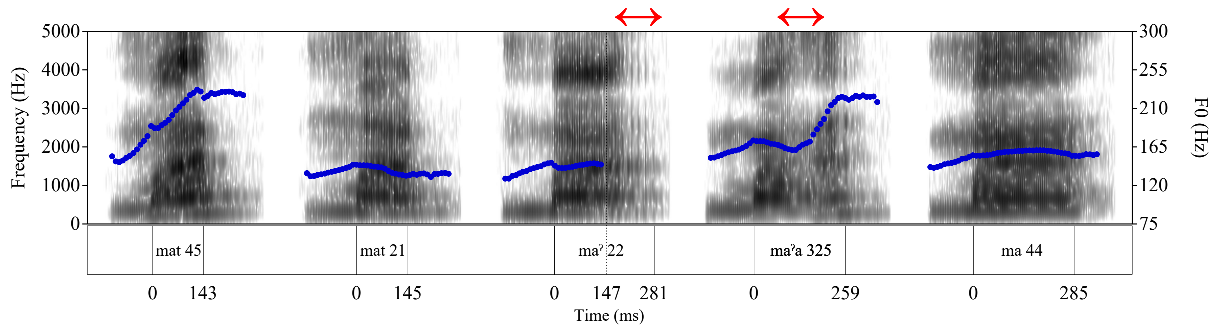


Figure 1.3. Spectrogram of checked D1 (45), checked D2 (21), B2 (22), C2 (325) and A1 (44) in Vietnamese. The red arrows mark the glottalized portions. The blue lines represent the F0.

The coda consonants of /mat 45/ and /mat 21/ are unreleased in the sample spectrograms (Figure 1.3). There is no visible creak in D1 and D2 tokens. Indeed, Michaud (2004), found that D1 and D2 tones have modal voice quality. B2 tone has a low F0 and irregular pulses in the end, suggesting a glottalized voice quality. C2 tone has low F0 and decreased amplitude in the middle of the vowel, suggesting a glottalized vowel quality in the middle. The modal tone A1 has modal voice quality throughout the vowel. In terms of the duration, the "checked" tones D1 (/mat 45/) and D2 (/mat 21/) have shorter duration than the "unchecked" glottalized tone B2 (/ma^{ʔ} 22/) and C2 /ma^{ʔ}a 325/, and the modal tone A1 (/ma 44/) .

The B2, D1, and D2 tones in Vietnamese all have obstruent-like closure at the end of the vowel. The B2 tone ends in glottalization, whereas the D1 and D2 tones are closed by oral stops. Tones B2, D1, and D2 share similar phonetic characteristics – having an abrupt turnoff of the voicing, which is usually an indicator of a checked constituent. However, D1 and D2 are identified as "checked" tones in studies of Vietnamese (e.g., Kirby 2011; Michaud 2004), whereas B2 is not. This indicates that whether a tone is regarded as "checked" also depends on its phonological structure. The B2 tone is borne by open syllables and nasal-closed syllables, whereas D1 and D2 tones are borne by oral stop-closed syllables (Michaud 2004). The B2 tone also has different origin from the D1 and D2 tone. The B2 tone is derived from words ending in /-ʔ/ in Proto-Viet-Muong, whereas D1 and D2 tones are derived from words ending in /-p,

-t, -k/ (Haudricourt 1954). This indicates that the term “checked” in Vietnamese is reserved for syllables that are closed by oral stops, and the tones borne by those syllables.

The phonetic and phonological properties of “checked” syllables and tones in Vietnamese are comparable to those in Meixian Hakka. The duration of “checked” syllable is shorter than “unchecked” syllables. The voice quality of vowels in “checked” syllables can be modal before the oral closure of the coda. “Checked” is used to refer to a syllable type and the tones that are associated with that syllable.

The “unchecked” B2 (22) tone in Vietnamese can be further compared with the “checked” vowel in Isthmus Zapotec. The B2 tone in Vietnamese and the checked vowel in Isthums Zapotec both have glottalization at the end of the vowel, and have a shorter duration (in terms of the modal voicing portion) than the modal tone or vowel. However, grammars of Vietnamese have not identified the B2 tone as being “checked.” This further indicates that Isthmus Zapotec and Vietnamese have different phonological definitions of “checked” constituents.

1.4 “Checked tone” in Burmese

Similar to Meixian Hakka and Vietnamese, Burmese is reported to have “checked” syllables and a “checked” tone (Gruber 2011). Checked syllables are those that are closed by a glottal stop. “Checked” syllables can only bear one tone, and that tone is named as a “checked” tone. The “checked” tone is also called “killed tone” in some studies (Watkins 2001; Green 2015). There are three other tones in Burmese – high, low, and creaky, which can be borne by open and nasal-closed syllables. Minimal pairs of “checked”, creaky, low, and high tones are in (4). The spectrograms of the tokens for “checked”, creaky, and high tones are in Figure 1.4.

(4)	“checked”	/maʔ 51/	“March”
	creaky	/ma̰ 51/	“female”
	high	/ma 44/	“towering”
	low	/ma 22/	“hard”
	(Watkins 2001, p. 293)		

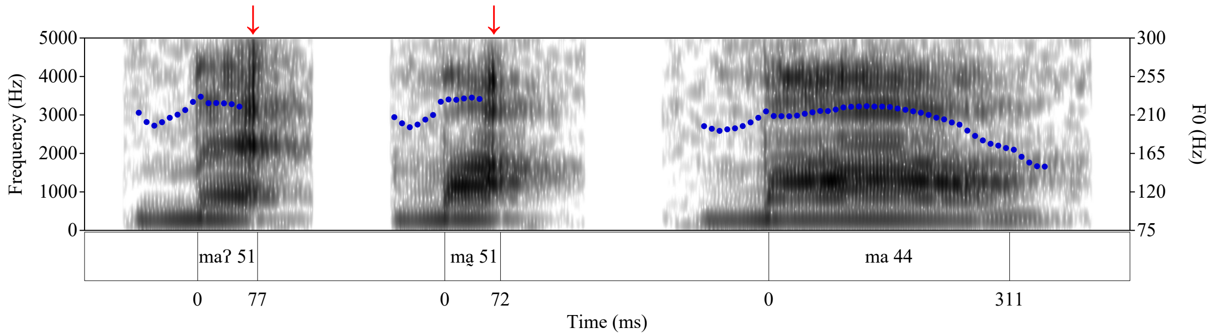


Figure 1.4. Spectrogram of "checked" tone (51), creaky tone (51), and high tone (44) in Burmese. The red arrows mark the glottalized portions. The blue lines represent the F0.

As shown in Figure 1.4, the F0, voice quality, and duration of "checked" and creaky tones are near-identical. "Checked" (/maʔ 51/) and creaky tones (/ma̰ 51/) both have falling F0 contour, glottalization at the end of the vowel, and shorter duration than the modal tone. Despite their similarity in the above three dimensions, "checked" tone and creaky tone can be distinguished by their vowel quality. In the spectrograms in Figure 1.4, [a] in /maʔ 51/ has higher F2 than [a] in /ma̰ 51/ and /ma 44/, suggesting that [a] in /maʔ 51/ is fronter than in open syllables. The vowel quality of /a/ in the /maʔ 51/ is similar to [æ].

The difference in vowel quality between checked syllable (CVʔ) and open syllable is systematic. In Burmese, CVʔ and CVN carry the same set of vowels (except for [ɛ], which is allowed in CVʔ but not CVN), and open syllables carry a different set of vowels, as shown in (5). CVʔ and CVN allow centralized vowels and diphthongs, whereas open syllables allow peripheral monophthongs. Although studies report that CVʔ and CV share [a] and [ɛ] (Watkins 2000, 2001; Gruber 2011), the formant values of [a] and [ɛ] differ between CVʔ and CV. As suggested by Watkins 2000, the [a] in CVʔ is higher and fronter than the [a] in open syllables in the F1–F2 space. The [ɛ] in [Cɛʔ] is lower and backer than [ɛ] in open syllables in the F1–F2 space. Thus, it might be more accurate to transcribe the [a] in CVʔ as [ʌ], and [ɛ] in CVN and CVʔ as [æ]. I did not find quantitative analysis of the vowel quality of [a] in CVN syllables. Based on the individual audio sample of /mã/ in Watkins 2001, the quality of /a/ resembles [æ].

More samples of /a/ from more CVN words are needed to make an accurate description of the vowel quality /a/ in nasal-closed syllables. In general, vowels in by CVʔ and CVN are more centralized than vowels in open syllables.

(5)	CV/CV̰	CVN/CV̰N	CVʔ
	[i]	[jɪN]	[jɪʔ]
	[e]	[eɪN]	[eɪʔ]
	[ɛ]		[ɛʔ] ([æʔ][a])
	[a]	[aN] ([æʔ][b])	[aʔ] ([ʌN][a])
		[aɪN]	[aɪʔ]
	[ɔ]	[aʊN]	[aʊʔ]
	[o]	[oʊN]	[oʊʔ]
	[u]	[ʊN]	[ʊʔ]

(Gruber 2011, p. 21)

[a]Revised transcription based on the formant values measured by Watkins (2000).

[b]Revised transcription based on the the audio sample of /mã/ "plaster" in Watkins 2001.

Glottal stop has been treated as a segment by studies of Burmese (e.g., Green 2015; Gruber 2011; Watkins 2001). The fact that CVʔ syllables share the same set of vowels with CVN syllables, but differ from open syllables in vowel quality, suggests that ʔ in CVʔ is a segment rather than a suprasegment. Only if ʔ is treated as a discrete segment, can CVʔ be treated as a closed syllable and behave the same as sonorant-closed syllables in vowel configuration. Moreover, the fact that CV syllables with the creaky tone do not differ in vowel quality from CV syllables with high or low tones suggests that the glottalization in creaky tone is a suprasegment rather than a segmental glottal stop. As a result, although we see vowel-final glottalization in both "checked" and creaky tones, the two tones differ in their phonological structure. Phonetically, the glottalization in the "checked" tone is the realization of a phonological glottal stop segment, whereas the glottalization in creaky tone is the realization of creaky phonation. We can formalize syllables with "checked" tones as /CVʔ/, and syllables with creaky tone as /CV^{ʔ}/, where the glottalization is a suprasegmental feature of the vowel.

Compared with Meixian Hakka and Vietnamese, the "checked" syllables and tones in Burmese have similar phonological structure but different in some phonetic properties. In these three languages, "checked" refers to obstruent-closed syllables and to the tone that is associated with such syllables. Phonetically, "checked" syllable in Burmese is characterized by final glottalization, whereas the "checked" syllables in Meixian Hakka (when the oral coda is released) and Vietnamese have modal-like quality. Such differences in voice quality could be caused by different codas in these languages. Burmese has glottal stop as the coda in "checked" syllables, which is expected to be realized as glottalization on the vowel (Garellek et al. 2021). Meixian Hakka and Vietnamese have oral stops as codas, which do not necessarily cause glottalization on the vowel.

We can also compare the "checked" and creaky tones in Burmese with the "checked" phonation in Isthmus Zapotec. Although the "checked" and creaky tones are both phonetically similar to the "checked" phonation in Isthmus Zapotec, only the "checked" tone is called "checked" in Burmese studies. This, again, indicates that the phonological definition of "checked" is different between Isthmus Zapotec and Burmese. For Isthmus Zapotec, we identify the final-glottalized phonation as "checked", whereas for Burmese, we identify the glottal stop-closed syllables and the tone borne by such syllables as "checked," but not those in which the vowel is glottalized.

1.5 Checked phonation, syllables, and tones

In the above examples from five languages, we see similarities and disparities in the phonetic and phonological properties of "checked" constituents across languages. In Table 1.1, I summarize the phonetic properties, the phonological structure of checked constituents, and the opposition of unchecked constituents in the languages.

Table 1.1. Phonetic properties and phonological structures of “checked” constituents in the literature

	Phonetic properties	Phonological structure		Phonological Opposition	
Isthmus Zapotec	short final-glottalized	Checked phonation	Vˀ	Modal phonation	V
				Laryngealized phonation	V̰
Hakka	short, modal or final-glottalized	Checked syllable	Vp,t,k	Unchecked syllable	V, VN
		Checked tone	55, 31	Unchecked tone	33, 53, 31, 11
Vietnamese	short modal	Checked syllable	Vp,t,k	Unchecked syllable	V, VN
		Checked tone	45, 21	Unchecked tone	44, 32, 24, 22, 412, 325
Burmese	short final-glottalized	Checked syllable	Vʔ	Unchecked syllable	V, VN
		Checked tone	Falling	Unchecked tone	High, Low, Creaky (falling)

Based on the summary of the “checked” constituents from the five languages in Table 1.1, we see that the shared phonetic characteristics of constituents that are identified as “checked” are short duration and having a stop-like closure at the end of the vowel. The closure can be either glottal or oral. Thus, I propose that from a phonetic standpoint, “checked” refers to an abrupt offset of voicing. “Abrupt offset of voicing” means that the amplitude of the vowel decreases suddenly. The opposition of “abrupt offset of voicing” is that in open syllables or syllables closed by sonorants, the amplitude of the vowel either trails off gradually (in utterance-final position) or does not has significant drop (in utterance-medial position). The abrupt offset of voicing is usually realized by having a glottal or oral stop closure at the end of the voicing. This phonetic property can be inferred from other names that are given checked constituents – e.g., “killed tone” (Burmese: Watkins 2001); “cut vowel” (Mitla Zapotec: Stubblefield and Hollenbach 1991; Tabaa Zapotec: Earl 2011).

Phonologically, examples in Table 1.1 illustrate that there are two different types of “checked” constituents. The first is a phonation type. Checked phonation has late-phased glottalization on vowels. It is in opposition with another type of glottalized phonation – laryngealized phonation, produced as mid-phased glottalization on vowels. The second is a restriction between syllable type and tone type. Checked syllables and checked tones always go together, such that a "checked syllable" that is distinct from an unchecked syllable will always be associated with a "checked" tone distinct from unchecked tones, and vice versa. Checked syllables are syllables that are closed by voiceless stops. The voiceless stop can be either oral or glottal. Checked syllables bear a different set of tones from unchecked syllables. Tones borne by checked syllables are called "checked tones."

It is necessary to distinguish what we mean by phonetically vs. phonologically "checked" because they are not equivalent. Not every phonetically checked constituent is phonologically checked. The criterion for determining whether a phonetically checked constituent is phonologically meaningful is whether the phonetically checked constituent behaves differently from the phonetically unchecked constituents in phonological processes. If the phonetically checked constituent is in contrastive opposition with phonetically unchecked constituent, and there is no existing feature or category that accounts for such a contrast, then it is necessary to use the term "checked" to distinguish the checked phonological group from the unchecked one. The phonation with late-phased glottalization in Isthmus Zapotec (Pickett, Villalobos, and Marlett 2010) meets this criterion. It contrasts with mid-phased glottalization. At the same time, there is no existing phonological category that distinguishes late-phased glottalization phonation from mid-phased glottalization phonation. It is necessary to label the late-phased glottalization phonation as the "checked" phonation, and establish it as a distinct phonological category.

An example of syllables that are phonetically checked, but which do not qualify as being phonologically checked, is that of syllables closed by voiceless stops in English (e.g. *sip*, *sit*, *sick*). In English, vowels in syllables closed by voiceless stops are shorter in duration than those in open syllables and syllables closed by voiced stops (Santen 1992, Chen 1970). Phonetically, this is similar to the checked syllables in Hakka, Vietnamese, and Burmese described earlier. However, I do not define those stop-closed syllables as checked syllables in English. The difference between English voiceless stop-closed syllables vs. Hakka, Vietnamese, and Burmese checked syllables is that the former do not differ from other syllable types in terms of phonological behaviors[3], whereas the latter differ from open syllables and sonorant-closed syllables in the languages in terms of tone distribution. It would therefore be unnecessary to assign a different name to voiceless stop-closed syllables in English; we can simply call /Vp, -t, -k/ in English "closed" syllables, on a par with all other closed syllables.

Why are checked phonation and checked syllables/tones regarded as distinct phonological constituents in this dissertation? How does one determine whether a language has a checked phonation or a checked syllable/tone? The phonological necessity and definition of checked will be discussed in Chapters 2 and 3. In Chapter 2, I discuss why it is necessary to identify a checked phonation type in certain Zapotec languages, and what the criteria are of determining whether a language has checked phonation. In Chapter 3, I discuss why it is necessary to identify checked syllables and checked tones in certain Chinese languages, and what the criteria are for determining whether a language has checked syllable and tone.

Given the phonological definition of checked constituents, the next question is, what are the phonetic properties of checked constituents? In this dissertation, I provide empirical evidence of checked syllables and tones in Xiapu Min, an under-documented variety of Min spoken in Fujian, China. Xiapu Min, like Hakka, Vietanmese, and Burmese, also has checked syllables and checked tones. Chapter 4 describes the acoustic properties of F0, voice quality, duration, and vowel quality of checked syllables and tones in Xiapu Min in production. Chapter 5 describes the acoustics properties of checked syllables and tones when they are said to be neutralized with unchecked syllables and tones in sandhi form. The results of the production studies will enrich the existing phonetic description of checked syllables and tones with original data from languages that have not been acoustically described or quantitatively analyzed before.

As Table 1.1 shows, checked constituents are usually short, and sometimes end in glottalization. Languages that have checked tones might also have a distinction between the checked

3. Vowels in syllables with voiced stop coda have longer duration than those in syllables with voiceless stop coda. However, duration is not contrastive in English. Thus, I do not regard the duration difference between syllables with voiceless stop coda and syllables with voiced stop coda as phonological differences between those two types of syllables

tones and unchecked tones in terms of F0. Such differences also shows up in Xiapu Min checked syllables and tones. The multidimensionality of the cues in production begs the question: do listeners use all the cues found in the production of checked tones when they are identifying checked tones in listening? Chapter 6 explores the function of each cue in perceiving checked syllables and tones in Xiapu Min in citation form using resynthesized stimuli. Chapter 7 explores whether the listeners can still differentiate checked tones from unchecked ones when they are phonologically neutralized to the same surface tone in sandhi form. If listeners can differentiate them, what cues do they use to perceive checkedness in neutralized forms? Few studies have investigated the perceptual cues for phonologically checked constituents, and even fewer studies identify the cues listeners use when checked and unchecked syllables are phonologically neutralized. The perception study results provided in this study will fill those research gaps and further our understandings of the phonetics of checked syllables and tones.

Chapter 2

Survey of checkedness in Zapotec languages

Zapotec languages are characterized as having checked vowels, in which "checked" refers to a phonation type with creaky voice realized at the end of the vowel. In Zapotec, the phasing of glottalization in the vowel is frequently found to be contrastive (e.g., Isthmus (Pickett, Villalobos, and Marlett 2010), Choapan (Lyman and Lyman 1977), Texmelucan Speck 1978b Zapotec). As Garellek et al. (2021) suggest, across languages, glottalization can be phased at the beginning, in the middle, at the end, or during the entire span of a vowel. I refer to those four phases as early-, middle-, late-phased, and in-phase glottalization, and schematize the gestural relation between vowel and glottal constriction in those four phases in Figure 2.1.

The naming of the phonations with different phasings of glottalization vary across Zapotec languages. Late-phased glottalization has been referred to as "checked," "glottalized," or "cut." Mid-phased glottalization has been referred to as "rearticulated," "interrupted," "broken," or "laryngealized." In-phase glottalization has been referred to as "creaky," or "laryngealized." No language in the survey has been reported to have early-phased glottalization (thought it occurs in other languages, such as Jalapa Mazatec; Garellek and Keating 2011). The names for mid-phased and in-phase glottalization sometimes overlap, because these realizations are frequently in free variation with each other. In this dissertation, I name late-phased glottalization as "checked," middle-phased glottalization as "rearticulated," and in-phase glottalization as “creaky.” In addition, when the phasing of the glottalization is not specified, or when changing the phasing of glottalization does not make a difference in the meaning, I refer to that phonation as “glottalized.” The phonetic symbol for each phasing type (adapted from Garellek et al. 2021), and the relation between the different glottalized phonation types is illustrated in Figure 2.1.

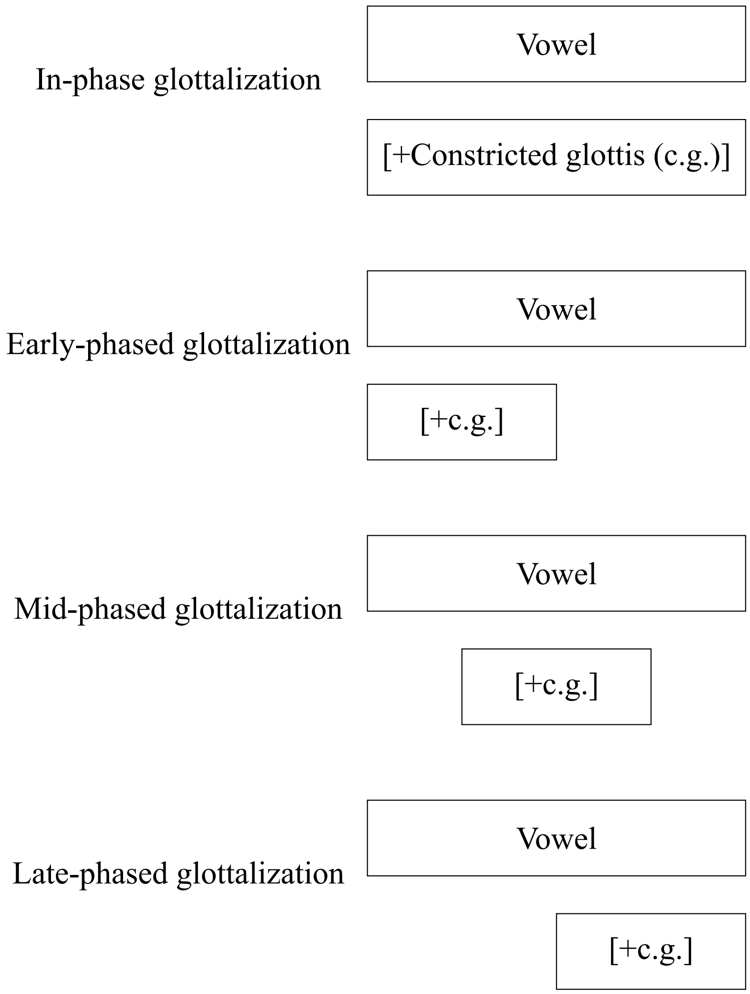


Figure 2.1. Phrasing of glottalization in early-, middle-, late-phased, and in-phase glottalizations.

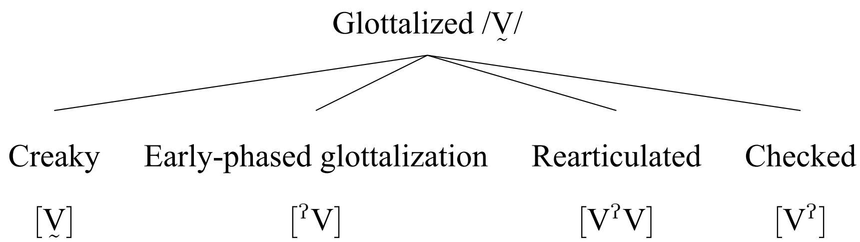


Figure 2.2. Names and symbols for, and relations between different phases of glottalization.

As shown in Figures 2.1 and 2.2, the definition of checked phonation is late-phased glottalization in vowels. Now let's return to the question in the introduction: is checked phonation a distinct phonological constituent from other possible constituents, namely short duration (phonological length), glottalized phonation, and tone? The fundamental difference lies in the distinction between glottalized phonation and checked phonation. By the definition given in this dissertation, glottalized phonation refers to any phonation that involves glottalization; the phasing of glottalization (relative to the vowel's other gestures) is not specified. Yet for checked phonation, the phasing of glottalization is specified as being late-phased, occurring at the end of the vowel. Thus, checked phonation is a subcategory of glottalized phonation, just like other three subcategories of glottalized phonations (creaky, early-phased, rearticulated). Each of these has different phasing of glottalization relative to the vowels. Thus, the term "checked" phonation contains information distinct from the more generic term – "glottalized" phonation, and its sister terms – "creaky", "early-glottalized", and "rearticulated" phonations. "Checked" needs to be treated as a distinct phonological category, one which involves the phonological feature (or gesture) [+constricted glottis] and a late-phased temporal restriction on vowels (Figure 2.1). Such temporal restrictions are phonologically necessary to account for the fact that many Zapotec languages have contrastive phasing of gltotalization. Although the choice of which precise phonological representation best captures contrastive phasing is not the focus here, it suffices to say that phonological theories that allow for subsegmental contrasts, like Articulatory Phonology (Browman and Goldstein 1986) or Q theory (Shih and Inkelas 2019) can better model these phonetic differences than phonological models of representation with no subsegmental structure.

In order to investigate the phonological and phonetic properties of checked phonation in Zapotec languages, I conduct a survey of 20 Zapotec languages (Table 2.1). The rest of this section contains three parts. Section 2.1 discusses the criteria for determining whether a given language has checked phonation. Section 2.3 illustrates, for languages with checked phonation, how the tone system and the checked phonation interact. Section 2.2 discusses for the Zapotec languages in the survey, whether the checked phonation is constrastive with another phonation type. Section 2.4 summarizes the phonetic feature of checked vowels reported in the literature.

Table 2.1. Languages in the survey of Zapotec languages (Oaxaca, Mexico). The glottocode and ISO 639-3 are retrieved from https://glottolog.org/.

Family	Zapotec language	Glottocode	ISO 639-3	Source
Central core	Tilquiapan	tilq1235	zts	Merrill 2008
Central core	San Francisco Ozolotepec	ozol1235	zao	Córdova 2009 Heise 2003 Leander 2008
Central core	San Juan Mixtepec	sanj1288		Nelson 2004
Central core	Santa Catarina Quioquitani	quio1242		Zurita Sánchez et al. 2008
Central core	Ayoquesco	ayoq1235	zaf	MacLaury 1970
Central core	Santa Ana del Valle			Esposito 2003, 2010
Central core	Teotitlán del Valle	sanj1284	zab	Uchihara and Gutiérrez 2019, 2020
Central core	Isthmus	isth1244	zai	Bueno Holle 2016 Marlett and Pickett 1987 Pickett et al. 2010
Central core	Guienagati	sant1463		Benn 2016, 2021
Central core	Mitla	mitl1236	zaw	Stubblefield and Hollenbach 1991
Central core	San Lucas Quiaviní			Chávez-Peón 2010, 2011
Core	Amatlán	amat1238	zpo	Riggs 2011
Core	Texmelucan	texm1235	zpz	Speck 1978a, 1978b, 1984
Northern core	Cajonos	cajo1238	zad	Nellis and Hollenbach 1980
Northern core	Choapan	choa1237	zpc	Lyman and Lyman 1977 Oliva-Juarez et al. 2014
Northern core	Yalálag	yala1267	zpu	Avelino 2004, 2016
Northern core	Betaza			Crowhurst et al. 2016 Teodocio Olivares 2009
Northern core	Zoogocho	zoog1239		Sonnenschein 2004
Northern core	Tabaa	taba1268	zat	Earl 2011
Northern core	Yatzachi	yatz1235	zav	Borroff 2007

2.1 Whether Vʔ is checked phonation: the status of glottal stop

When we observe a language with vowels produced with final (i.e., late-phased) glottalization, we can hypothesize that such vowels are checked vowels. In order to evaluate whether the language indeed has checked phonation, we need to first determine the phonological status of glottal stop in phonetic instances of Vʔ. In addition, because the opposition between V^{ʔ} and V^{ʔ}V means that checked phonation may be phonologically contrastive, this section will also discuss the phonological status of the glottal stop in the phonetic instances of VʔV as well. If what are transcribed as glottal stops in Vʔ and VʔV are considered to be suprasegments (i.e., realization of the feature of [+constricted glottis] at the end and in the middle of vowels), the glottal gesture in Vʔ and VʔV can be considered part of the phonetic implementation of phonation type. Otherwise, if the glottal stops in Vʔ and VʔV are analyzed as segments, then we should not treat Vʔ and VʔV as having glottalized phonations at the phonological level. We should not propose a checked or rearticulated phonation for Vʔ or VʔV. The goal of this section is to summarize the evidence provided in the literature on Zapotec (along with some neighbouring languages) for determining the phonological status of glottal stop, to evaluate the effectiveness of each piece of evidence, and to propose criteria for determining whether glottal stop is a segment or not. This section starts with evidence for and against glottal stop as a suprasegment in Vʔ structures, followed by evidence for and against glottal stop as a suprasegment in VʔV structures, and ends with a interim summary of the evidence that this dissertation advocates for determining whether glottal stop is a segment or not. Most of the phonological evidence supports the view that Vʔ and VʔV refer to phonation types, where the glottalization is suprasegmental.

2.1.1 Phonetic glottal stop analyzed as a suprasegment in Vʔ (/V^{ʔ}/)

This section summarizes the reasons for treating the glottal stop in the phonetic sequence of Vʔ as a suprasegment (i.e., features of [+constricted glottis] realized at the late phase of vowels: /V^{ʔ}/), as stated in the existing literature on Zapotecan languages. The following five reasons have been found:

1. Vʔ behaves the same as open syllables (CV) and differently from closed syllables (CVC) in phonological processes;

2. Vʔ behaves the same as VʔV or V̰ but differently from syllables with modal vowels (CV

or CVC);

3. Glottal stop has a limited distribution in the language;

4. Analyzing glottal stop as a suprasegment simplifies the syllable structure of the language;

5. Phonetic evidence

Reasons 1–4 are phonological in nature whereas Reason 5 is phonetic. Among the types of phonological evidence, Reasons 1 and 2 are more definitive than Reasons 3 and 4 because there are counterarguments to the latter two reasons. Phonological evidence is prioritized over phonetic evidence because the decision of the *phonological* status of glottal stop should be supported by phonological evidence. The phonetic realization of glottal stop has been found to be highly varied (Garellek et al. 2021). Arguments based on phonetic realization of glottal stop thus cannot definitively determine the phonological status of glottal stop. I will present and analyze the five reasons listed above in this section with examples from specific grammars.

An example of Reason 1 – Vʔ behaving the same as open syllable (CV) but differently from closed syllable (CVC) in phonological procedures – is from Texmelucan Zapotec. Texmelucan Zapotec is reported to have modal, glottalized <V’>[1] and laryngealized <VV> vowels (Speck 1978b). Texmelucan Zapotec requires morpheme contraction for certain morphemes when the roots that they modify are open syllables. For example, the third person possessive morpheme /-mi/ becomes a suffix to the root when the root is an open syllable (6a), but becomes a separate word when the root is a closed syllable (6b). When the root has a glottalized /Vʔ/ or a laryngealized /V̰/ vowel, the contraction of /-mi/ into suffix [-m] applies. Example (6) indicates that the glottal stop in /Vʔ/ does not function the same as other coda consonants in closed syllables. /Vʔ/ functions the same as open syllables. Glottal stop in /Vʔ/ structure is thus most likely to be a suprasegment that is part of the vowel, rather than a discrete segment.

1. The symbols in angle brackets <> are the orthographic representations of the phones or words used in the literature.

An alternative analysis to the phenomenon in Example (6) is that the glottal stop in /Vʔ/ is still a segment and the different suffixation patterns between syllables closed by obstruents and syllables closed by glottal stop is due to phonotactics. Third person possessive morpheme /-mi/ is realized as a suffix /-m/ in /juʔ/ but as a separate morpheme in closed syllable /ʃab/ because /bm/ is not allowed as a complex coda in the language. However, roots in Texmelucan Zapotec do not otherwise have complex codas. If glottal stop is a segment, we will have to claim that /ʔm/ is the only possible complex coda in Texmelucan Zapotec. This explanation is less favorable than the explanation that glottal stop function as a phonation feature because it requires more complicated phonotactic rules for the language.

(6)		Root	Gloss	Underlying form	Surface form	Gloss
	a.	/lo/	"face"	/lo-mi/	[lom]	"her face"
	b.	/ʃab/	"clothes"	/ʃab-mi/	[ʃab mi]	"her clothes"
	c.	/juʔ/	"house"	/juʔ-mi/	[juʔm]	"her house"
	d.	/ja̰/	"hand"	/ja̰-mi/	[ja̰m]	"her hand"

(Speck 1978b, p. 14)

Another example for Vʔ patterning the same as open syllables but differently from closed syllables is from Guienagati Zapotec. Benn (2021) proposed that Guienagati Zapotec has three types of phonation: modal (V), checked (Vʔ), and rearticulated (VʔV). The person markers in Guienagati Zapotec are conditioned by the syllable structure. When the root is a closed syllable, the marker for first person singular is an enclitic /-àʔ/. When the root is an open syllable, the first person marker is realized as vowel nasalization on the nucleus of the root (7a). When the root is of Vʔ structure, the first person marker is also realized as nasalization on the vowel (7b). When the root ends in a consonant, the first person marker is an affix /=à/ attached to the root (7c). The marker of first person in Vʔ structured syllables is the same as in open syllables but different from closed syllables, suggesting that Vʔ is treated as an open syllable in Guienagati Zapotec. Glottal stop in Vʔ is accordingly treated as a suprasegmental feature of the vowel.

(7)	Root	Gloss	1st possessed	Gloss
a.	/lè/	“name”	/da-lȅ/	“my name”
b.	/mbalbjèʔ/	“shrimp”	/mbalbjȅʔ/	“my shrimp”
c.	/dūj/	“blackbird”	/da-dūj=à/	“my blackbird”

(Benn 2021, p. 54)

Reason 2 – Vʔ behaving the same as syllables with non-modal vowels (e.g. VʔV or V̰) but differently from syllables with modal vowels (CV or CVC) – is based on the rationale that if Vʔ is grouped with another non-modal phonation and is opposed to modal phonation, Vʔ should also be considered as a phonation type, which requires ʔ to be a suprasegmental phonatory feature realized on the vowel. An example for Reason 2 comes from Guienagati Zapotec (Benn 2016). There are three types of codas in Guienagati Zapotec: geminate (Cː), voiced (C), and voiceless (C̥). The distribution of coda is conditioned by phonation type and tone. Syllables with modal phonation and low or mid tone require voiced and geminate consonants as codas. Syllables with modal phonation and high tone allow voiceless consonants as codas. Syllables with checked or rearticulated phonation allow voiced or voiceless consonants as coda (Table 2.2). According to Benn 2016, checked and rearticulated phonation do not have contrastive tones. The distribution of coda shows that phonation type has an effect on the type of codas allowed. Non-modal phonations share the same coda distribution and behave differently from modal phonation. If glottal stop is analyzed as a segment, it will complicate the phonotactic rules of the language.

Table 2.2. Distribution of coda by phonation and tone in Guienagati Zapotec (Benn 2016, p. 126)

Phonation	Tone	Coda
Modal	Low	C vs. Cː
	Mid	C vs. Cː
	High	C̥
Checked		C vs. C̥
Rearticulated		C vs. C̥

Several studies appeal to Reason 3 – Glottal stop has a limited distribution in the language – when determining that glottal stop is a suprasegment in the language. Such studies include Sonnenschein 2004: Zoogocho Zapotec, Chávez-Peón 2010: Quiaviní Zapotec, Nellis and Hollenbach 1980: Cajonos Zapotec, Nelson 2004: San Juan Mixtepec Zapotec. They propose that glottal stop should be treated as a suprasegment because it only occurs in coda position but never in onset position. This reason is not as convincing as Reasons 1 and 2, because a restricted distribution is not sufficient to disqualify a unit as a segment. Consonants have restricted distributions in many languages. For example, /ʔ/ and /h/ in Chamicuro only occur in coda position but not in onset position (Parker 2001). However, /ʔ/ and /h/ are both segments in Chamicuro. Chamicuro has contrastive vowel length. Long vowels can only occur in open syllables. Only short vowels are found before /ʔ/ and /h/, indicating syllables ending in /ʔ/ and /h/ are more likely to be closed. In addition, Chamicuro does not allow complex codas, and /ʔ/ and /h/ do not cooccur with other consonants in the coda position of the same syllable. The distributional patterns of vowel and coda consonant in syllables ending in /ʔ/ and /h/ in Chamicuro indicate that /ʔ/ and /h/ are segments in the language. Another well-known counterexample is that /ŋ/ in English only occurs in coda position. However, this does not disqualify /ŋ/ from being a segment, given that there are oppositions between CVŋ vs. CVm vs. CVn (e.g. hang /hæŋ/ vs. ham /hæm/; sung /sʌŋ/ vs. sun /sʌn/).

Reason 4 given in the literature for determining glottal stop in Vʔ structure as a suprasegment is that analyzing glottal stop as a suprasegment simplifies the syllable structure of the language. For example, Pickett, Villalobos, and Marlett (2010) suggest that, in Isthmus Zapotec, glottal stop in CVʔ should be regarded as a suprasegment because otherwise it will be the only possible coda in the language. For the simplicity of the syllable structure of the language, they propose treating glottal stop in CVʔ as a suprasegment, and analyze all syllables in Isthmus Zapotec as open syllables. However, this is not definitive evidence for glottal stop as a suprasegment in the language. There are cases where laryngeals (/h/ and /ʔ/) are the most frequent codas or the only codas of the language. Parker (2001) found that in a corpus with 700 lexical items of Chamicuro (Southern Maipuran: Chamicuro-Morike, cham1318, ccc), 93.6% of closed syllables have /ʔ/ or /h/ as their coda. Similarly, in Burmese citation forms, /ʔ/ is the only possible coda (Watkins 2001; Gruber 2011). Those examples show that it is possible to have glottal stop as the only coda in the language. For the case of Isthmus Zapotec, although analyzing glottal stop as a suprasegment feature of the vowel is plausible, it does not rule out the possibility that glottal stop is a segment either.

Reason 5 for treating glottal stop in Vʔ as a suprasegment is phonetic evidence. Teodocio Olivares (2009) suggests that there is a contrast between Vʔ and Vˀ in Betaza Zapotec. The ʔ in Vʔ is a segment that is realized with an aspirated release of the glottal closure ([Vʔʰ]), whereas the ʔ in Vˀ is a suprasegment realized without an aspirated release. The contrast between Vʔ and Vˀ provided by Teodocio Olivares (2009, p. 27), [láʔʰ] vs. [lâˀ], is not a minimal opposition between aspirated glottal stop and unaspirated glottal stop, because they also differ in tone. It is therefore unclear whether it is the high vs. falling tone, or the aspiration of the glottal stop, or both are the contrasting unit(s) of this near-minimal pair. It is also unclear for words that are transcribed as [ʔʰ] and [ˀ], how consistently the glottal stop is aspirated or unreleased. The realization of glottal stop is highly varied within and across languages (Garellek et al. 2021), and the sound can be realized as strongly as a full glottal stop with aspirated closure, or as weakly as modal voicing with a decrease in amplitude. Glottalized phonation can also be realized as weakly as modal voicing with a decrease in amplitude, or as strongly as a full glottal stop (e.g. Yalálag Zapotec: Avelino 2004). Using phonetic evidence to decide the phonological structure of glottal stop is thus not favored.

2.1.2 Phonetic glottal stop analyzed as a segment in Vʔ (/V+ʔ/)

Next, I review the arguments that support glottal stop as a segment in Vʔ structure (i.e., phonologically /V + ʔ/). Studies on Zapotec have listed five reasons:

1. Vʔ has the same stress assignment as VC, but different from V;

2. Vʔ only occur in open syllables, whereas V and VʔV can occur in closed syllables;

3. Glottal stop occurs in consonant clusters as other consonants in the language;

4. Glottal stop can occur in both onset and coda positions;

5. Phonetic evidence.

Reasons 1 and 2 are from the same language: Cajonos Zapotec. In Cajonos Zapotec polysyllabic words, stress falls on the final syllable for the words ending in Vʔ or VC (e.g., /ʧíˈbēʔ/ "crab"; /xósˈtís/ "authority"), and falls on the penultimate syllable for the words ending in V or VʔV (e.g. /ˈʒwágé/ "pitcher"; /ˈbèjéʔe/ "ice") (Nellis and Hollenbach 1980, p. 100). The alignment in the stress assignment suggest that Vʔ share similar phonological structure as VC, whereas VʔV shares similar phonological structure as V. This suggests that the glottal stop in Vʔ is more likely to be segmental, whereas the glottal stop in VʔV is more likely to be suprasegmental. In addition, in Cajonos Zapotec, the Vʔ structure is never followed by another consonant (*VʔC), whereas VʔV and V can both be followed by a consonant coda (e.g. /nì/ "here"; /jìʧ/ "paper"; /dàʔa/ "palm mat"; /s̬àʔag/ "town errand boy"). As Nellis and Hollenbach (1980) proposed, it is more reasonable and common to propose that the glottal stop in /Vʔ/ is a segment and there is a rule that complex coda is not allowed in the language (i.e., *ʔC) than to propose that the glottal stop in /Vʔ/ is a suprasegment feature of vowel and there is a rule that vowels with checked phonation are not allowed in closed syllables (i.e., *V^{ʔ}C). Reasons 1 and 2 together suggest that the glottal stop in Vʔ in Cajonos Zapotec is more likely to be segmental.

Reason 3 for determining that glottal stop in Vʔ is a segment – Glottal stop occurs in consonant cluster as other consonants in the language – is from Yalálag Zapotec (Avelino 2004). Avelino (2004) treated glottal stop in the Vʔ structure in Yalálag Zapotec as a consonant. Glottal stop can occur alone as the coda of a syllable, or can occur with [χ] (allophone of /g/) as a consonant cluster in the coda position. Other consonants in Yalálag Zapotec behave similarly such that they can be either a standalone coda, or form a complex coda with [χ] (e.g., [nχ], [pχ], [tχ]), which is the only possible segment for the second segment of a complex coda when the first segment is not a glottal stop. Neverthless, Avelino (2004)'s arguments can be challenged for two reasons. First, treating glottal stop in Vʔ as a suprasegment can also explain this phenomenon. If glottal stop is a suprasegment, Vʔ is an open syllable and Vʔχ is a syllable closed by a single coda. Second, while [χ] is the only permissible second segment for consonants other than glottal stop to form a complex coda, glottal stop can form a complex coda with any consonant when it is the first segment (Avelino 2004, p. 146). The possible complex codas with glottal stop as the first segment that were listed in the examples include ʔj, ʔʧ, ʔt, ʔb, ʔz, ʔχ, and ʔn. Examples for complex coda in Yalálag Zapotec are in Table 2.3. If glottal stop in Vʔ is analyzed as a segment, more phonotactic rules are required for the language, such that glottal stop can occur with any consonant to form a complex coda whereas other consonants can only occur with [χ] to form a complex coda. Alternatively, if glottal stop in Vʔ is a suprasegment, no further phonotactic rules for complex coda are needed, since the consonant after Vʔ becomes the single coda of the syllable. Based on the above two reasons, I propose that the reason given by Avelino (2004) for glottal stop in Vʔ as a segment is not strong enough to support its status as unequivocably segmental.

Table 2.3. Consonant clusters in coda in Yalálag Zapotec (Avelino 2004, p. 148).

Coda consonant cluster	Word	Gloss
nχ	[βðinχ]	"edge of a cliff"
pχ	[jâpχ]	"chayote (type of squash)"
tχ	[zìtχ]	"deep"
ʧχ	[bèʧχ]	"crow"
ʔχ	[wʧéʔχ]	"to tie"
ʔj	[za̰ʔj]	"cheek"
ʔʧ	[bèʔʧ]	"louse"
ʔt	[wèʔt]	"to sell"
ʔb	[wràʔb]	"to count (numbers)"
ʔz	[waʧèʔz]	"to split"
ʔs	[wàʔs]	"to chew"
ʔn	[wʒéʔn]	"to bend over"

Reason 4 – Glottal stop occurs in both onset and coda position – comes from a non-Zapotecan language – Yucatec Maya (Core: Yucatec-Lacandon, yuca1254, yua). According to Frazier (2013), Yucatec Maya has four tones, three of which are modal whereas one of which is glottalized: Short (short, no tone, modal voice: V), Low tone (long, low tone, modal vowel: V̀V), High tone (long, high tone, modal voice: V́V), and Glottalized (long, high tone, creaky voice: V́V̰). Frazier 2013 provided examples of words with phonetic transcription that begins and ends in a glottal stop: [ʔam] "spider"; [siʔ] "firewood". Words with glottal stop following a long vowel (low toned, high toned, and glottalized vowels) are rare. The examples of long glottalized vowel followed by a glottal stop given in Frazier (2013) are either allophonic variation of /ɓ/ (e.g.[táa̰ɓ] ∼[táa̰ʔ] "salt") or words after morphological affixation (e.g. /tiʔ -iʔ/ →[tíḭʔ] "there-LOCATIVE"). However, the data provided by Frazier (2013) is not sufficient to determine whether the glottal stop occurring in the onset and the coda positions in Yucatec Mayan is a segment or not. First, the glottal stop observed in onset could be the result of glottal stop insertion before word-initial vowels (Hayes 2011). Epenthetic glottal stop before word-initial vowels has been widely observed. For example, in Tektitek Maya (Core: Mamean, tekt1235, ttc), glottal stop is inserted before word-intial vowels. When the vowel-initial word has a consonant prefix, the glottal stop insertion is prohibited (e.g. [ʔaɓ̥aχ] "stone" vs. [w-aːɓ̥aχ] "my stone"; Bennett 2016, p. 477). In Ñumí Mixtec, glottal stop is consistently inserted before word-initial vowels regardless of utterance position (Gittlen and Marlett 1985). An overview of languages with glottal insertion can be found in Garellek 2013. Positive evidence for glottal stop in onset position being phonemeic is to see whether there is opposition of glottal onset and zero onset. One example is Tongan (Austronesian: Polynesian, tong1325, ton) (e.g. /ˈʔena/ "their" vs. /ˈena/ "withdraw oneself"), where this constrast remains even in hiatus and in utterance-initial position (Garellek and Tabain 2020). In terms of coda, the presence of a glottal stop at the end of a syllable in the phonetic transcriptions of the words does not necessarily speak to the phonological status of the glottal stop. Only if there is phonological evidence that syllables closed by glottal stop functions the same as syllables closed by other consonants but differently from open syllables, can we make the conclusion that ʔ in Vʔ functions as a segment. Frazier (2013) provided an example where /ɓ/ becomes [ʔ] in word-final position: [táa̰ɓ] ∼ [táa̰ʔ] "salt". This example indicates that [ʔ] is an allophone of [ɓ]. However, Frazier (2013) did not provide examples of /ʔ/ functioning as a segment in the coda position in the *underlying* form. In summary, even though an glottal stop occurs in the onset and coda position in the phonetic transcription of the examples for Yucatec Maya, it is unclear whether those glottal stops are underlyingly segments or not.

Reason 5 uses phonetic evidence to determine glottal stop in Vʔ as a segment . For example, Avelino (2004) found that the duration of the vowels in Vʔ structure is as short as the vowels in syllables closed by voiceless stops. The vowels in Vʔ and VC are significantly shorter than the modal and laryngealized vowels in open syllables. Avelino (2004) treated this as evidence that glottal stop in Vʔ patterns the same as the coda consonant in VC syllables. However, this is not the only possible analysis for the short duration of the vowels in Vʔ syllables. Counterarguments can be found in Chávez-Peón 2010, 2011 using data from Quiaviní Zapotec. Quiaviní Zapotec has Vʔ and VʔV syllables. Chávez-Peón (2010, 2011) analyzed the glottal stop in Vʔ as part of the vowel. The duration of Vʔ should include both the voicing portion of the vowel and the glottal closure and the voiceless echo vowel following the release of the glottal closure. Although the voicing portion of Vʔ is shorter than the voicing portion of VʔV, the overall duration of Vʔ with the glottal closure included is comparable to the overall durations of VʔV and long vowels. Thus, the short duration of the voicing portion in Vʔ can be explained by either treating the glottal stop as a segment so that the vowel is shorter before voiceless coda; or treating the glottal stop as part of the vowel so that the voicing portion is just part of the vowel and the overall duration of the vowel $V^{ʔ}$ is actually long. The short duration of the voicing portion in Vʔ consequently is not a definitive evidence for glottal stop as a segment in Vʔ.

The other studies appeal to the strong degree of glottalization in the realization of ʔ in Vʔ as evidence of glottal stop being a segment. Borroff (2007) concluded that the glottal stop in Vʔ in Yatzachi Zapotec is a segment because there are examples where ʔ in Vʔ is realized as a full glottal stop (p. 49) in careful speech. Since careful speech is assumed to closely approximate underlying forms, the full glottal closure realization of Vʔ indicates that glottal stop is a segment rather than a glottalized phonation. As mentioned in Section 2.1.2, Teodocio Olivares (2009) suggested that ʔ realized with an aspirated release of the glottal closure is a segment whereas those realized without aspirated release is suprasegment. Esposito (2003) found that the voice quality (represented by measures of spectral tilt) of the vowels in Vʔ or V̰ʔ does not differ from the voice quality of the modal or “creaky” vowels in open syllables (V and V̰), and concluded that glottal stop in Vʔ is a segment. However, the word list used by Esposito (2003) is based on phonetic transcription. I did not encounter minimal pairs between V and Vʔ or between V̰ and V̰ʔ. Thus, it is unclear whether the “creaky” phonation and the glottal stop transcribed in the word list are phonemic or not. To summarize, I do not regard the above arguments based on the phonetic realization of glottal stop or the vowel before the glottal stop as definitive evidence for glottal stop as a segment.

2.1.3 Phonetic glottal stop analyzed as a suprasegment in VʔV (/V^{ʔ}V/)

Determining the phonological status of glottal stop in VʔV is relatively simpler than in Vʔ because we can resort to the number of syllables in VʔV and the behavior of the suprasegmental features related to syllable structure, such as tone and stress. I summarize seven reasons provided by the existing literature for treating glottal stop in VʔV as a suprasegment (i.e., phonologically /V^{ʔ}V/):

1. VʔV carries the same (or a smaller) tonal inventory as single syllables;
2. Native speakers of the language treat VʔV as one syllable;
3. VʔV carries the same vowel inventory as monosyllabic syllables;
4. VʔV is counted as a single syllable in stress assignment;
5. The vowel quality before and after the ʔ in VʔV is always homogeneous;
6. Analyzing ʔ in VʔV as a suprasegment simplifies the morphophonological structure of the language;
7. Phonetic evidence.

Reason 1 uses the tone pattern in VʔV to decide whether ʔ is a segment or not. To illustrate this reasoning, let's assume a language with four lexical tones and with the syllable as the tone-bearing unit (TBU). If ʔ is a suprasegment, VʔV contains one syllable and can only carry the tones that are allowed in other monosyllables. If ʔ is a segment in the language, VʔV contains two syllables and should be able to carry up to 16 possible tone combinations (4 * 4 tones). None of the languages that are covered by the current survey display a tone pattern of the latter type. For example, Avelino (2016) concludes that in Yalálag Zapotec, VʔV is a single segment rather than two syllables because VʔV only contain one TBU. VʔV and modal V carry the same set of tones: high, low, and falling. Chávez-Peón (2011) points out that, in Quiaviní Zapotec, VʔV structure does not have a larger tonal inventory than syllables with modal vowel as nucleus. In fact, modal nucleus can carry high, low, falling, rising tones, whereas VʔV structure can only carry low and falling tones. Nelson (2004) also stated that glottal stop in San Juan Mixtepec Zapotec is not a segment, but is instead laryngealization on the vowel because VʔV has only one TBU. Modal vowels can occur with high, low, falling, and rising tones. Glottalized vowels (Vʔ or VʔV) carry low, falling, or rising tones in the surface forms.

Reason 2 for glottal stop in VʔV being a segment is native speakers' intuition. Avelino (2016) found that native speakers of Yalálag Zapotec used a single whistled note, tap, and hum for words with VʔV vowel, as they did for monosyllables with modal vowels. Similarly, Chávez-Peón (2010) asked native speakers of Quiaviní Zapotec to separate words by syllable. Native speakers grouped VʔV as one syllable. When asked to whistle or clap once per syllable, they produced a single whistled note or clap for VʔV. Another example is based on word game on a non-Zapotecan language. Kekchi (Q'eqchi') (Core: Greater Quichean, kekc1242, kek) is a Mayan language mainly spoken in Guatemala (Campbell 1974). Campbell (1974) performed a word game "jerigonza" with the native speakers of Kekchi. The rule of jerigonza is to add /p/ and a copy of the vowel after each vowel. For example, CV_1CV_2 will be converted into $CV_1pV_1CV_2pV_2$. The word /ʧaʔax/ "difficult" was converted into either [ʧaʔapax] or [ʧapaʔapax]. Speakers who converted /ʧaʔax/ to [ʧaʔapax] treated /aʔa/ as one vowel, whereas those who converted it to [ʧapaʔapax] treated /aʔa/ as two syllables (whether ʔ functions as a coda or an onset is unclear). The result of jerigonza for the word /ʧaʔax/ "difficult" suggests that there is individual variation in whether to treat the glottal stop in VʔV as segment for Kekchi speakers.

Reason 3 appeals to the vowel quality of VʔV compared with that of monosyllables. In Quiaviní Zapotec, vowels before and after the glottal stop in VʔV either have the same vowel quality, or have the same quality as the diphthongs occurring with modal phonation (e.g. [gìʔà] "market", [gagjèʔì] "around", [ɾtíʔà] "gathers", [ɾzéʔìɲ] "get capricious") (Chávez-Peón 2011, pp. 19-20). The comparable vowel quality between VʔV structures and monosyllables with modal vowel suggests that VʔV is also a monosyllable with a single nucleus. Consequently, ʔ in VʔV cannot be a segment. Otherwise, we expect more combinations of monophthongs beyond the existing diphthongs in the vowel inventory of the language.

Reason 4 – VʔV is counted as a single syllable in stress assignment – has already been discussed in Section 2.1.2 when discussing the stress pattern in Cajonos Zapotec. In Cajonos Zapotec, polysyllabic words ending in an open syllable have the stress on the penultimate syllable. VʔV is counted as one single syllable in the stress assignment. For example, in word /bèjéʔe/ "ice", the stress falls on the syllable /bè/. Consequently, the glottal stop in VʔV cannot be a segment.

Reason 5 – the vowel quality before and after the ʔ in VʔV is always homogeneous (also known as translaryngeal harmony) – is similar to Reason 3. Avelino (2016) concludes that the glottal stop in VʔV is a suprasegment in Yalálag Zapotec and Ocotepec Mixe (Mixe-Zoque: Juquila Mixe, ocot1241) because the vowel before and after glottal stop is always identical. The homogeneity of vowel before and after the glottal stop indicates VʔV as a single segment rather than a sequence of three segments. Avelino (2016) also uses the lack of translaryngeal homogeneity as the evidence that the glottal stop in Vʔ structure is a segment. Avelino (2016) finds that having different vowel qualities is allowed when two Vʔ syllables are juxtaposed to each other (i.e. V_1ʔ.V_2ʔ; e.g. [ʒàʧâjràʔóʔ] "You are suffering"), which suggests that glottal stop is a segment that functions as other consonants. Translaryngeal homogeneity has also been observed in Cajonos Zapotec (Nellis and Hollenbach 1980). Cajonos Zapotec does not allow two different vowels to occur consecutively (*V_1V_2). When a sequence of vowels is created due to morphological process, the second vowel is either dropped or becomes a semivowel. The same rule applies to laryngealized vowels (VʔV). Different vowels before and after laryngealized vowels have not been observed (*V_1ʔV_2). Examples are in (8). However, if there is a non-glottal stop consonant separating two different vowels, the vowel quality of both vowels is preserved (V_1CV_2 or V_1ʔV_1CV_2). In other words, ʔ is not a sufficient barrier to avoid a hiatus environment.

(8)

	Underlying form	Gloss	Surface form	Gloss
a.	/ʂ-tò-àʔ/	"of-string-my"	[ʂtwàʔ]	"my string"
b.	/ʂ-á-éʔ/	"of-father-his"	[ʂéʔ]	"his father"
c.	/ʐél·éʔe-àʔ/	"of-banana-my"	[ʐél·âʔa]	"my banana"
d.	/dàʔa-bīʔi/		[dàʔaˈbīʔi]	"this morning"

(Nellis and Hollenbach 1980, p. 99, 102)

However, it is debatable whether translaryngeal homogeneity can be definitive evidence for glottal stop being a suprasegment. There are alternative analyses for translaryngeal homogeneity. Duanmu (1994) and Borroff (2007) both provided analyses of glottal stop as a segment in translaryngeal homogeneity. Duanmu (1994) proposed that if vowel assimilation is blocked by consonant but is transparent to glottal stop (e.g. Garo: /aʔ-i/ →[aʔ-a]; /ak-i/ →[ak-i]), it could be due to the glottal stop lacking an oral feature. Glottal stops do not have an oral place of articulation in the supralaryngeal cavity. Thus, there can be a rule in the language that vowel assimilation is prohibited only when two vowels are separated by a consonant with an oral feature. As a result, glottal stop is transparent to vowel assimilation because it lacks oral feature, whereas consonants with a supralaryngeal place of articulation blocks vowel assimilation. Borroff (2007) explained translaryngeal harmony using Articulatory Phonology and Optimality Theory. Within articulatory phonology, Gafos (2002) specified that there are five abstract gestural landmarks for every segment: ONSET, TARGET, C-CENTER, RELEASE, and OFFSET (see Example (15) on page 14 in Borroff (2007) for the schematized temporal relation among the landmarks). ONSET is "the onset of movement towards the target gesture". OFFSET is "the end of all movement associated with the gesture." Borroff (2007) proposed a revised ONSET constraint, which requires that "for every vocalic gesture the ONSET landmark of some consonantal gesture should be synchronous with the ONSET landmark of that vocalic gesture. (p. 98)" According to Borroff (2007), glottal stop is underspecified for its ONSET and OFFSET gesture because it does not have a supralaryngeal gesture. As a result, glottal stop is unable to have its ONSET to be aligned with the ONSET of the vowel. $V_1ʔV_2$ consequently violates the ONSET constraint. If the ONSET constraint ranks high, V_1 and V_2 have to coalesce into the same vowel quality, or one of the vowels has to drop, in order to make V_1ʔV_2 surface as one syllable in the output and avoid violating the ONSET constraint. Both Duanmu (1994) and Borroff (2007)'s analyses appeal to the lack of supralaryngeal feature of glottal stop. Phonetically, it is also plausible that the lack of oral constriction between two vowels motivates the vowel quality of the preceding vowel to be carried onto the following vowel, or the preceding vowel is anticipatorily assimilated into the vowel quality of the following vowel. The formal phonological analyses by Duanmu (1994) and Borroff (2007) and the articulatory characteristics of glottal stop together suggest that translaryngeal homogeneity (V_1ʔV_1) is not a definitive reason for determining glottal stop as a suprasegment. The alternative analysis that glottal stop is a discrete segment but behaves differently from other supralaryngeal consonants is also plausible.

Reason 6 – analyzing ʔ in VʔV as a suprasegment simplifies the morphophonological structure of the language – is comparable to Reason 4 when determining glottal stop as a suprasegment in Vʔ in Section 2.1.1. In Ocotepec Mixe (Avelino 2016), Quiaviní Zapotec (Chávez-Peón 2010), and San Juan Mixtepec Zapotec (Nelson 2004), native roots are primarily monosyllabic. If ʔ is a segment, all words containing VʔV will be disyllabic, which complicates the overall syllable structure in the languages. However, this reason alone is not definitive, because it does not appeals to the phonological properties of VʔV per se. We need strong evidence, such as VʔV behaving as other monosyllables in phonological processes, to conclude that the glottal stop in VʔV is indeed suprasegmental.

Reason 7 is based on phonetic evidence. Frazier (2013) proposed that the traditionally referred "rearticulated" vowels that are denoted as /VʔV/ should be reanalyzed as /VV̰/ in Yucatec Maya. Glottal stop should be a suprasegment because it is most frequently realized as creaky voice. As I stated in the previous sections, I maintain that phonetic evidence alone cannot decide the phonological status of a segment.

2.1.4 Phonetic glottal stop analyzed as a segment in VʔV (/V+ʔ+V/)

There is only one study in my survey of Zapotec languages suggesting that ʔ in VʔV is a segment (i.e., phonologically /V + ʔ + V/). Borroff (2007) suggested that ʔ in Yatzachi Zapotec is a segment because it is realized with full glottal stop closure in careful speech. Borroff (2007) used the same reason when identifying ʔ in Vʔ as a segment. As I have suggested in Sections 2.1.1–2.1.3, phonetic evidence alone is not sufficient evidence to decide the phonological status of glottal stop.

2.1.5 Interim summary

In Table 2.4, I summarize the criteria for glottal stop being a segment and for glottal stop being a suprasegmental feature of vowel, and the languages that meet each criterion.

Table 2.4. Criteria for glottal stop being a segment and suprasegment in Vʔ and VʔV with examples from Zapotec languages

ʔ status	Language	Criteria	Sources
suprasegmental /Vˀ/	Texmelucan	Vʔ pairs with V but not VC	Speck 1978b
	Guienagati	Vʔ pairs with V but not VC	Benn 2021
		Vʔ pairs with rearticulated vowel but differently from modal vowel	Benn 2016
segmental /Vʔ/	Cajonos	Vʔ has same stress distribution as VC but different from V	Nellis and Hollenbach 1980
		Vʔ cannot be followed by consonant coda but V and VʔV can	Nellis and Hollenbach 1980
suprasegmental /VˀV/	Yalálag	VʔV is treated as one TBU rather two: the tonal inventory of VʔV is not larger than V	Avelino 2016
		Native speakers treat VʔV as one syllable (using methods such as whistling, tapping, humming)	Avelino 2016
	Quiaviní	VʔV is treated as one TBU rather two: the tonal inventory of VʔV is not larger than V	Chávez-Peón 2011
		VʔV is treated as one TBU rather two: the vowel inventory of VʔV is not larger than V	Chávez-Peón 2011
	San Juan Mixtepec	VʔV is treated as one TBU rather two: the tonal inventory of VʔV is not larger than V	Nelson 2004
	Cajonos	VʔV is counted as a single syllable in stress assignment	Nellis and Hollenbach 1980
segmental /VʔV/	N/A	N/A	N/A

The principle of the criteria is to use evidence that can confirm that glottal stop behaves the same as a consonant or as a vowel phonologically. The phonological status of glottal stop needs to be evaluated separately in Vʔ and VʔV environments. Glottal stop being a suprasegment in one structure does not imply that the glottal stop in other structures is also suprasegmental. In the criteria above, we see that convincing examples have been found for glottal stop as a suprasegment in both Vʔ and VʔV environments. There is one example for glottal stop as a segment in Vʔ (Cajonos Zapotec). This suggests that, for Zapotec languages, it might be typologically more common to have glottal stop as a suprasegment than as a segment.

2.2 Is checked phonation contrastive with other phonations?

Once we have evidence that [Vʔ] is a phonation type (i.e., /Vˀ/), the following questions is: is the checked phonation is a contrastive phonation type in the language? As illustrated in Figure 2.2, checked phonation is a type of glottalized phonation. The difference between checked phonation and glottalized phonation is that the former has late-phased glottalization [Vˀ], whereas the latter refers is an abstract term with the phasing of glottalization unspecified /V̰/. Thus, when changing late-phased glottalization to middle-phased makes a difference in word meaning, it is necessary to specify that there is a contrastive checked phonation (and a contrastive rearticulated phonation) in the language. The term "glottalized" alone is not sufficient to capture the phasing contrast. Otherwise, checked phonation is allophonic in a language. For example, if a language has free variation between [Vˀ] and [VˀV], then the checked phonation and rearticulated phonation are allophones of each other.

In this section, I will review the phonation system of the 20 Zapotec languages in the survey, and describe whether each language has checked phonation, and whether the checked phonation is allophonic or contrastive using the criteria in Sections 2.1 and 2.2. For languages that do not discuss the segmental status of glottal stop, I assume that glottal stops in Vʔ and VʔV are suprasegmental, based on the general trend in the Zapotec languages. Future studies should look for evidence to determine the phonological structure of glottal stop. I categorize the 20 languages in the survey by the number of contrastive glottalized phonation types in the language. There are seven languages that have a single glottalized phonation type, twelve languages that have two glottalized phonation types, and one language that has three glottalized phonation types.

2.2.1 Languages with one contrastive glottalized phonation type

For languages with only one type of glottalized phonation (Table 2.5), checked cannot be a contrastive phonation of another glottalized phonation. There are four languages with checked phonation as allophonic to rearticulated phonation, two languages with no checked phonation, and one language in which the checked phonation status is undetermined.

In Tilquiapan, San Francisco Ozolotepec, San Juan Mixtepec, and Ayoquesco Zapotec, the checked phonation is allophonic to rearticulated phonation. The alternation is in free variation in Tilquiapan and San Francisco Ozolotepec Zapotec, an is governed by phonological rules in San Juan Mixtepec and Choapan Zapotec.

The glottalized vowels in Tilquiapan (Merrill 2008) and the laryngealized vowels in San Francisco Ozolotepec Zapotec (Leander 2008) are realized as [V^{ʔ}] or [V^{ʔ}V] in free variation.

In San Juan Mixtepec Zapotec, glottalized vowels are realized as either [V^{ʔ}] before fortis consonants (e.g., [ʃò^{ʔ}p] "six"), or as [V^{ʔ}V] before lenis consonants or in pre-pausal position (e.g., [mè^{ʔ}e͡dz] "jaguar"; [dò^{ʔ}o] "rope"; Nelson 2004, p. 7).

In Ayoquesco Zapotec, two glottalized phonation types are identified: glottalized vowels [V^{ʔ}] and interrupted vowels [V^{ʔ}V], and they are allophonic to each other (MacLaury 1970). /V^{ʔ}V/ ("interrupted vowel") becomes [V^{ʔ}] ("glottalized vowel") when it is in a monomorphemic word and functions as a prefix to a noun (e.g. /tā^{ʔ}a-bìʔin/ [tā^{ʔ}bì^{ʔ}in] "*muchachos* (young men)").

Amatlán Zapotec (Riggs 2011) and Santa Catarina Quioquitani Zapotec (Zurita Sánchez, Ward, and Marlett 2008) are reported to have laryngealized vowels. The laryngealized vowels are realized as in-phase glottalization [V̰] or rearticulated glottalization [V^{ʔ}V] in free variation. No checked phonation is reported in these two languages.

Cajonos Zapotec is reported to have "checked" phonation in contrast with laryngealized phonation. "Checked" vowels [Vʔ] are realized as vowel followed by a glottal stop in the production. Laryngealized phonation [VʔV] is produced with glottalization or full glottal stop in the middle of the vowels. "Checked" vowels are in contrast with laryngealized vowels (/zíʔ/

“hurt” vs. /zíʔi/ “heavy”). However, as introduced in Section 2.1.2 Reasons 1 (i.e., Vʔ has the same stress assignment as VC) and 2 (i.e, Vʔ cannot be followed by consonants), the glottal stop in [Vʔ] in Cajonos Zapotec is more likely to be segmental. In contrast, as introduced in Section 2.1.3 Reason 4 (i.e., VʔV is counted as a single syllable in stress assignment), the glottal stop in [VʔV] is more likely to be suprasegmental. Thus, although the author proposes that Cajonos Zapotec has contrastive “checked” and laryngealized phonations, the “checked” vowels [Vʔ] are in fact not a phonation type. It is a syllable closed with a glottal stop. The only non-modal phonation in the language is the laryngealized/rearticulated phonation /Vˀ V/.

Lastly, in Santa Ana del Valle Zapotec, the existence of checked phonation is undetermined. Esposito (2003, 2010) have measured the mean H1–H2 and H1–A3 values of the “creaky” vowels, and found that “creaky” vowels have more glottal constrictions than modal vowels. However, the phasing of the glottalization in the vowels was not discussed in these studies. Further analysis is required to measure the change of voice quality during the time course of vowels to determine whether the creaky voice is realized with late-phased glottalization in the language. However, even if there is phonetic evidence of creaky voice being produced with late-phased glottalization, the checked phonation is not contrastive in the language, because there is no evidence showing that the phasing of glottalization is contrastive. The status of checked phonation for languages with one type of glottalized phonation is summarized in Table 2.5.

Table 2.5. Checked phonation status in languages with one type of glottalized phonation

Zapotec language	Has checked phonation?	Checked phonation contrastive?	Name of checked phonation in sources	Name of other non-modal phonation in sources
Tilquiapan	Y	N	Glottalized	NA
San Francisco Ozolotepec	Y	N	Laryngealized	NA
San Juan Mixtepec	Y	N	Glottalized	NA
Ayoquesco	Y	N	Glottalized	Interrupted
Amatlán	N	NA	NA	Laryngealized
Santa Catarina Quioquitani	N	NA	NA	Laryngealized
Cajonos	N	NA	NA	Laryngealized
Santa Ana del Valle	NA	NA	NA	Creaky, Breathy

2.2.2 Languages with two contrastive glottalized phonation types

In the survey, I found nine languages that contrast two types of glottalized phonations. All of them have checked phonation in contrast with another glottalized phonation. The other glottalized phonation has either in-phase glottalization or middle-phased glottalization. I summarize such languages and their phonation system in Table 2.6. In the table, I list the names given in the literature for the checked phonation in the language. When the reference only provided non-IPA transcription, I put the notation in angle brackets. I also include the names and transcription of other phonations in the languages. The /Vˀ/ represents late-phased glottalization. The /VˀV/ phonation represents middle-phased or in-phase glottalization. A minimal pair between the checked phonation and the other glottalized phonation is provided for each language.

Choapan Zapotec is proposed to have "checked" vowels and "laryngeal" vowels. The "checked" vowels are [Vˀ]. "Laryngeal" vowels are realized as [Vˀ] when followed by another vowel, or as [VˀV] in other environments (9). For the so-called "laryngeal" vowel before another vowel, it is debatable whether $[V_1{}^{ʔ}V_2]$ are two vowels, or is a single vowel $[\widehat{V_1{}^{ʔ}V_2}]$, as in Quiaviní Zapotec where the rearticulated vowels can also be diphthongs and the glottalization is between the first and the second vowel in the diphthong. In order to determine whether $[V_1{}^{ʔ}V_2]$ is a checked vowel followed by an unchecked vowel, or is a rearticulated diphthong, we need to know whether $[V_1{}^{ʔ}V_2]$ bears one tone or two, and whether the vowel inventory of $[V_1{}^{ʔ}V_2]$ is larger than the inventory of diphthong in the language. If $[V_1{}^{ʔ}V_2]$ only bears a single tone, and coincide with the diphthong inventory in the language, the $[V_1{}^{ʔ}V_2]$ is more likely to be a diphthong. Despite the debate over the status of "laryngeal" vowel, there are near-minimal pairs between [V^{ʔ}] and [V^{ʔ}V] in the language: [le^{ʔ}e] "you.PL"; [ba^{ʔ}a] "flat"; [be^{ʔ}] "ice"; [-ba^{ʔ}] "3rd person animal."

(9) a. /lḛo/ [le^{ʔ}o] "wall of a house"
b. /lɛ̰/ [lɛ^{ʔ}ɛ] "side, stomach"
(Lyman and Lyman 1977, p. 142)

Note that Avelino (2004, 2016) treated ʔ in the Vʔ structure in Yalálag Zapotec as glottal stop, rather than a phonation type, as discussed in Section 2.1.2 Reasons 3 (consonant cluster pattern) and 5 (phonetic evidence). However, those two reasons are not definitive. Moreover, Avelino (2004, 2016) referred to Vʔ as "checked vowels." They compared the phonetic properties of Vʔ with rearticulated/laryngealized vowels, and also discussed the tonal distribution among modal vowel, rearticulated/laryngealized vowels, and checked vowels. It is phonetically and phonologically meaningful (in terms of tone) to compare checked vowels with rearticulated/laryngealized vowels in the language. For the reasons above, I put Yalálag Zapotec in the category of languages that have two contrastive glottalized phonation types.

Table 2.6. Checked phonation status in languages with two types of glottalized phonation. The symbols in angle brackets <> are the orthographic representation of the sound used in sources.

Zapotec language	Has checked phonation?	Checked phonation contrastive?	Name of checked phonation in sources	Name of other non-modal phonation in sources	Minimal pair
Teotitlán	Y	Y	glottalized	creaky [V̰]	[ruˀ] "mouth" [rṵ] "carry"
Isthmus	Y	Y	checked	laryngealized [V̰]	[gìˀ] "excrement" [ʒḭ] "nose"
Choapan	Y	Y	checked	laryngeal <V̲>	[-baˀ] "3rd person animal" [baˀa] "flat"
Yalálag	Y	Y	checked	rearticulated [VˀV]	[gaˀ] "green" [gaˀa] "basket"
Betaza	Y	Y	glottalized	laryngealized [VˀV]	[bèˀ] "wind" [béˀe] "this morning"
Texmelucan	Y	Y	glottalized	laryngealized <VV>	[zaˀ] "fresh corn" [zaˀa] "bean"
Guienagati	Y	Y	checked	rearticulated [V̰]	[baˀ] "throat" [baˀa] "cemetary"
Zoogocho	Y	Y	checked	creaky <VV>	[jaˀ] "reed" [jaˀa] "steam bath"
Tabaa	Y	Y	cut *cortada*	broken <V'V> *quebrada*	[laˀ] "my name" [laˀa] "beans"
Mitla	Y	Y	cut *cortada*	broken <VV> *quebrada* breathy <Vj> *aspirada*	[saˀ] "wedding" [saˀa] "good"

2.2.3 Language with three glottalized phonation types

In this survey, I find one language – San Lucas Quiaviní Zapotec – that is reported to have three types of glottalized phonation: "checked" [Vˀ] or [Vʔ$^{\text{V̥}}$], "rearticulated" [VˀV], and "creaky" [V̰] (Chávez-Peón 2010, 2011). "Checked" and "rearticulated" vowels are in complementary distribution. "Checked" vowels can only bear high tone, whereas "rearticulated" vowels can only bear low or falling tone. They are grouped together and called "interrupted" phonation. "Checked" and "rearticulated" vowels only occur in open syllables or syllables closed by lenis consonants. "Creaky" vowels can bear high, low, or falling tone, and can occur in open syllables or syllables closed by fortis or lenis consonants. The distribution of "checked", "rearticulated", and "creaky" phonation by syllable type and tone category is presented in Table 2.7. As a result, although "checked" and "rearticulated" phonations are in complementary distribution in terms of the tone, "checked" and "rearticulated" phonations can each be in contrast with the "creaky" phonation. Example (10) shows near-minimal pair between "creaky" and "checked" (a), minimal pair between "creaky" and "rearticulated" (b), and the complementary distribution of "checked" and "rearticulated" (c). The transcription for the underlying and surface forms of the examples are all from Chávez-Peón 2010.

Table 2.7. Distribution of "checked", "rearticulated", and "creaky" phonations by syllable type and tone category in Quiaviní Zapotec (Chávez-Peón 2010, 2011)

	Open			VC_{lenis}			VC_{fortis}		
	L	H	F	L	H	F	L	H	F
"Creaky"	✓	✓	✓	✓	✓	✓	✓	✓	✓
"Checked"	×	✓	×	×	✓	×	×	×	×
"Rearticulated"	✓	×	✓	✓	×	✓	×	×	×

(10)		Phonation	Underlying form	Surface form	Gloss
	a.	"Creaky"	/ʒḭm˙j ˥/	[ʒí̬mːj][a]	"basket"
		"Checked"	/ʒiˀ ˥/	[ʒíʔⁱ̥]	"cold"
	b.	"Creaky"	/ba̰ ˩/	[bà̰ː]	"tomb"
		"Rearticulated"	/baˀ ˩/	[bàˀà]	"eyeball"
	c.	"Checked"	/ɾgaˀ ˥/	[ɾgáʔᵃ̥]	"gets green again"
		"Rearticualted"	/ɾgaˀ ˩/	[ɾgàʔà]	"gets caught"

(Chávez-Peón 2010, pp. 212, 240–241)

[a] Chávez-Peón (2010) used /m˙/ to represent that the sonorant is a fortis consonant; [V̬] to represent that the vowel is realized with a tense voice.

The difference between "checked" and "rearticulated" phonation types is in the timing of the glottalization: "checked" phonation has strong glottalization at the end of vowels, whereas for rearticulated phonation, it is in the middle. Chávez-Peón (2011, p. 14) grouped "checked" and "rearticulated" together as "interrupted" phonation because "they use the same laryngeal mechanism: extreme glottalization (either glottal closure or very pronounced creakiness)." Also, "checked" vowels have an optional voiceless echo vowel after the glottal constriction ([Vʔ$^{\text{V̥}}$]), whereas "rearticulated" vowels have a mandatory voiced vowel after the glottal constriction ([V^{ʔ}V]).

The difference between "creaky" and "checked" is the degree of glottalization. "Checked" vowels are realized with a stronger degree of glottalization than "creaky" vowels. Both "checked" and "creaky" vowels have late-phased glottalization in vowels. "Creaky" vowels can be in either open syllables or closed syllables with either fortis or lenis coda, whereas "checked" vowels can only occur in open syllables or syllables closed by lenis consonants. "Checked" and "creaky" vowels both can bear high tone, whereas "creaky" vowels can also bear low and falling tone.

Although "creaky" and "checked" phonation can occur in the same environment – open syllables or syllables closed by lenis consonant in high tone – Chávez-Peón (2010, 2011) did not provide a minimal pair that contrasts "creaky" vs. "checked" phonations in either environment. At the same time, there has yet been reported a language that uses the degree of glottalization contrastively, independently of its phasing. Given the absence of minimal pair between "creaky" and "checked" phonation, here I propose an alternative analysis of the phonation system of San Lucas Quiaviní Zapotec: "Creaky" and "checked" phonations are not contrastive. They are allophonic variants of the same phonation type. We can thus rename "creaky" and "checked" phonation types as *checked* phonation and transcribe them as [V^{ʔ}], because they are characterized by late-phased glottalization in vowels. The checked vowels are realized as strong late-phased glottalization in some lexical items, but as weak late-phased glottalization in others. I do not find a rule that can explain the distribution of strong glottalization and weak glottalization by phonological conditions. It is likely that the strong/weak glottalization variation is lexical. If we group "creaky" and "checked" phonation as a single phonation – checked phonation, Quiaviní Zapotec consequently has only *two* contrastive glottalized phonation types: checked and rearticulated.

Based on this alternative analysis, the checked vowels in the language can bear high, low, and falling tone and occur in any open and closed syllables, whereas the rearticulated vowels in the language can bear low and falling tone and occur in open and lenis consonant-closed syllables. I present my alternative analysis for the phonation system in San Lucas Quiaviní Zapotec in Table 2.8. In the minimal pair provided in the table, the example for the checked phonation is from an example for the "creaky" phonation in Chávez-Peón 2010.

Table 2.8. Checked phonation status in San Lucas Quiaviní Zapotec

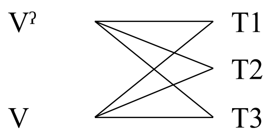
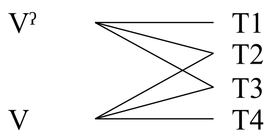
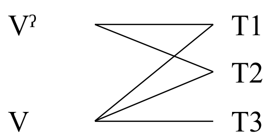
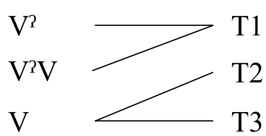
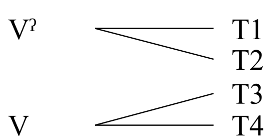
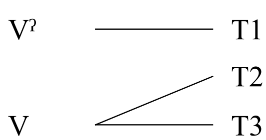
Zapotec language	Has checked phonation?	Checked phonation contrastive?	Name of checked phonation in sources	Name of other non-modal phonation in sources	Minimal pair
San Lucas Quiaviní	Y	Y	checked, creaky	rearticulated /Vˀ V/ breathy /V̤/	/bàˀ/ "tomb" /bàˀa/ "eyeball"

2.3 Interaction between tone and checked phonation

Sections 2.1 and 2.2 have discussed the criteria for checked phonation, and discussed the phonation contrasts in Zapotec languages. Most Zapotec languages are also tonal. The question for the current section is: how do tone and the checked phonation interact in Zapotec languages?

Kuang (2017) proposed two categories of interaction between phonation and tone. One is that languages can contrast different phonations independently of tone. The other is that tone and phonation covary with each other. The examples given for the former type of interaction are Jalapa Mazatec and Southern Yi. Tones and phonation types are fully crossed in Jalapa Mazatec (Garellek and Keating 2011) and Southern Yi (Kuang and Keating 2013). The examples given for the latter type are Cantonese (Yu and Lam 2014), Northern Vietnamese (Brunelle 2009), White Hmong (Garellek et al. 2013), Green Hmong (Andruski and Ratliff 2000), and Mandarin (Kuang 2017). Non-modal phonation is associated with different tone(s) in those languages from modal phonation. After reviewing the tone and phonation type interaction in Zapotec languages, I find that the two categories of tone–phonation configurations proposed in Kuang 2017 can be further divided into subcategories. For example, languages can contrast different phonations independently of tone, even though tones and phonation types are not fully crossed. In some Hmong and Zapotec languages, non-modal phonations have a smaller tonal inventory than the modal phonation (Kuang 2013). For languages whose phonation types covary with tones, there are cases where each non-modal phonation only has one designated tone, and there are also cases where non-modal phonations bear contrastive tones. In order to capture the full variation in phonation–tone interactions, I maintain that it is worth discussing the interaction between tone and phonation from two dimensions: whether the language contrasts phonation independently from tone, and whether the language contrasts tone independently from phonation. "Contrasting phonation independently from tone" means that there is minimal pair in the language where the tone is the same but the phonation is different (e.g. high-toned modal vowel vs. high-toned creaky vowel). "Contrasting tone independently from phonation" means that there is minimal pair in the language where the phonation is the same but the tone is different (e.g. high-toned modal vowel vs. low-toned modal vowel). Based on those two dimensions, I summarize six types of phonation-tone interaction. Since this dissertation focuses on the phonology of checked phonation, I will use checked (Vʔ) vs. modal (V) phonations as the sample contrast pair in the illustration. This model applies to any other phonation types by replacing Vʔ and V with any other phonation types found in the language. Note that it is possible that different phonation pairs in the same language belong to different tone-phonation interaction types.

Table 2.9. Types of tone–phonation interaction

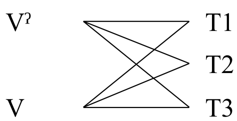
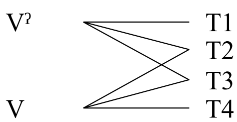
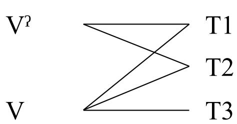
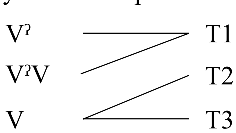
Tone-phonation interaction type		Contrast between checked phonation and modal phonation independently of tone	Contrast each tone pair independently of phonation
Type I: Fully-crossed		✓	✓
Type II: Overlapped		✓	✓
Type III: Subset		✓	✓
Type IV: Non-modal phonation types carry the same tone which is not carried by the modal phonation		✓($V^ʔ$ vs. $V^ʔV$)	×
Type V: No overlap–multiple tones		×	✓
Type VI: No overlap–one tone		×	×

Types I, II, and III all contrast phonation independently of tone, as well as tone independently of phonation. They differ from each other by the configuration between the phonation and the tone. Type I has phonation and tone fully crossed with each other. Checked phonation and other phonations may carry the same set of tones. Type II has the tones of checked phonation overlapping with the tones of other phonations. Checked vowels do not carry T4 whereas other phonations do not carry T1. Type III has the tones of checked vowel being a subset of the tones carried by other phonations. Languages belonging to Type I, II, and III have minimal pairs that contrast checked phonation with other phonations, with tone being held constant. They also have minimal pairs that contrast each tone with another tone, with phonation being held constant.

For a language to be of Type IV, it must have three phonation types. Here I use checked $V^{ʔ}$, rearticulated $V^{ʔ}V$, and modal V as an example. Checked and rearticulated vowels do not have contrastive tones, they only carry T1. Modal vowels have contrastive tones, but the tones carried by the modal vowels do not overlap with the tone carried by the non-modal phonations. For languages of Type IV, the checked phonation is indeed contrastive with another phonation – the rearticulated phonation, independently of tone. There is minimal pair of T1 checked vs. T1 rearticulated. However, T1 does not contrast with other tones independently of phonation. There is no minimal pair of T1 modal vs. T2/T3 modal, or T1 checked vs. T2/T3 checked. Type V have checked phonation and the other phonation carrying different set of tones. Checked phonation contrast multiple tones. Languages of Type V contrast tones independently of phonation. Each tone has minimal pairs where the phonation is held constant and the tone changes. On the other hand, there is no minimal pairs where the tone is held constant and the contrast is between checked and the other phonation. Thus, languages of Type V do not contrast checked phonation with other phonations independently of tone. The last type, Type VI, have the checked phonation associated with only one tone, and that tone does not overlap with the tones carried by the other phonation. In this case, tone and phonation covary with each other. Languages of Type VI do not contrast checked phonation with other phonations independently of tone. Neither do they contrast T1 with other tones independently of phonation.

In Table 2.10, I review the phonation and tonal inventories of Zapotec languages covered in the survey, and analyze in each language, which type the interaction between checked phonation and tone belongs to. For languages that do not have contrastive checked phonation, I analyze which type the interaction between glottalized phonation and tone belongs to. For several languages introduced in Section 2.2, I did not find description for their tone system in the literature, and thus excluded them from Table 2.10.

Table 2.10. Categorization of Zapotec languages by the interaction between checked phonation and tone. For languages without contrastive checked phonation, the interaction between the glottalized phonation and tone is presented. Those languages are marked with (V̰) in the Language column.

Interaction type	Zapotec language	V	$V^{ʔ}$	$V^{ʔ}V$	V̰	V̤
Type I: Fully-crossed 	Teotitlán	H M L F R	H M L F R		H M L F R	
	Isthmus	H L R	H L R		H L R	
	Choapan	H M L	H M L	NA[a]		
	Tilquiapan (V̰)	HL			HL	
	Santa Catarina Quioquitani (V̰)	H L R			H L R	
	Choapan (V̰)	H M L			H M L	
	Cajonos ($V^{ʔ}V$)	H M L F		H M L F		
Type II: Overlapped 	Ayoquesco (V̰)	H M L F R Extra L			H M L R Extra H Extra L	
Type III: Subset 	San Juan Mixtepec (V̰)	H L F R			L F R	
	San Lucas Quiaviní	H L F R	H L F	L F		L F
	Yalálag	H L F	H L	H L F		
	Betaza	H L F R	H L F	H L F R		
	Texmelucan	H M L F	H L	M L		
Type IV: Non-modal phonations carry the same tone which is not carried by the modal phonation 	Santa Ana del Valle (V̰)	H R			F	F

[a]No examples for [$V^{ʔ}V$] in high tone was found. However, it is unclear whether [$V^{ʔ}V$] never bears high tone or the examples in Lyman and Lyman 1977 did not include a token with [$V^{ʔ}V$] bearing high tone.

As Table 2.10 shows, the languages in the survey belong to Types I, II, III, or IV. Languages of Type I, II, and III all have checked phonation contrast with unchecked phonation independently of tone; and have tonal contrasts independent of phonation. One example was found for Type II interaction – Ayoquesco Zapotec, for which modal and non-modal phonation have overlapping tonal distributions. Moreover, each phonation type may bear a tone that the other cannot. Modal vowels do not bear the extra-high tone, whereas non-modal vowels do not bear the falling tone (MacLaury 1970).

One example was found for Type IV interaction – Santa Ana del Valle Zapotec, where the "creaky" phonation contrasts with the breathy phonation independently of tone, but the falling tones does not contrast with another tone independently of phonation. The "creaky" phonation in Santa Ana del Valle Zapotec is in contrast with breathy phonation, because they can both bear falling tone. However, the falling tone does contrast with other tones independently of phonation, because it is the only designated tone for non-modal phonation types. No examples have been found for Type V (checked and modal phonation bear different tones; checked phonation bears more than one tone.) or VI (checked and modal phonation bear different tones; checked phonation bears only one tone). In summary, all the Zapotec languages (that have description of the tonal system in the literature) have checked or glottalized phonation contrast with other phonation type independently of tone. I thus conclude that checked phonation in Zapotec languages tends to be independent of tone.

In addition, there are two languages for which it is unclear whether tone and phonation are independent or not. One is Guienagati Zapotec. Benn (2016) maintained that tone is not contrastive for Vʔ and VʔV vowels in Guienagati Zapotec. Phonetically, Vʔ is realized with a falling pitch and VʔV is realizaed with a mid-ranged pitch. The other case is Coatlan-Loxicha Zapotec (Beam de Azcona 2004). It is reported to have a glottal tone with a rising pitch contour and four types of realization of the glottalization based on the phonological structure of the TBU (whether the syllable coda is an obstruent; whether the coda is voiced; and whether the syllable is word-final), glottalization can be realized phonetically as checkedness or rearticulation on vowels, or with pre-/post-glottalization of codas). For Guienagati Zapotec and Coatlan-Loxicha

Zapotec, there are two possibilities of analyzing the relation between tone and phonation. One possibility is to treat the phonetic prosodic realization of the non-modal phonations in Guienagati Zapotec and Coatlan-Loxicha Zapotec as phonemic tone. Then we can claim that Guienagati Zapotec has a falling tone for checked vowels and a mid tone for rearticulated vowels. As a result, checked vowels have a dedicated tone because no other phonation types bear the falling tone. Guienagati Zapotec would then belong to the Type VI category (i.e., The language does not contrast phonation independently of tone. Neither does it contrast the falling tone independently of phonation.) For Coatlan-Loxicha Zapotec, we can claim that the glottal tone is the combination of glottalized phonation and a rising tone. Since modal vowels can also bear the Rising tone, the tonal inventory of glottalized phonation is a subset of the tonal inventory of modal phonation. Coatlan-Loxicha Zapotec would then be a Type II language.

The alternative analysis is to treat the non-modal phonation types in Guienagati Zapotec and Coatlan-Loxicha Zapotec as toneless: phonation is the contrastive feature, and the pitch associated with those non-modal phonation types is part of the phonetic realization of the phonation. This analysis is plausible because Beam de Azcona (2004) mentioned that, in Coatlan-Loxicha Zapotec, the pitch contour of the "glottal tone" is more variable than other tones. Changing the pitch contour of "glottal tone" does not affect the word meaning, and glottalization is the most stable feature. If we regard the non-modal phonations in Guienagati Zapotec and Coatlan-Loxicha Zapotec as toneless, we can conclude that tone and phonation in those two languages do not interact. The most direct way of knowing which analysis is more appropriate is through perceptual experiments, in which the cues of pitch and phonation are isolated, and presented to listeners of these languages. Then we will know whether phonation or pitch, or both matters more for the listeners to determine the identify of a word.

Besides the independent interaction between phonation types and tones on word-level, the independence between phonation and tone in Zapoteca languages is also observed in the grammatical transformations. In Teotitlán Zapotec, phonation and tone can perform grammatical functions independently of each other (Uchihara and Gutiérrez 2020). For some verb stems, there is an alternation between glottalized phonation and modal phonation between personless and first person forms. Example (11) shows that when the verb is not marked by person, /i/ and /a/ in the stem have glottalized phonation (ri-gḭ̄ːly, ri-gā̰ːyn), whereas when the verb is marked by a first-person subject, /i/ and /a/ are in modal phonation (ri-gīːly=á, ri-gāːny=á). Yet, the tone of the stems remains the same in the two examples. The alternation between glottalized vowels and modal vowels in Example (11) shows that the glottalized phonation in Teotitlán Zapotec has a grammatical function that is independent of the tone.

(11)		Personless		1ST person	
	a.	ri-gḭ̄ːly	"look for (it)"	ri-gīːly=á	"I look for (it)"
	b.	ri-gā̰ːyn	"dig"	ri-gāːny=á	"I dig"

Not only does the phonation alternation occurs independently of tone in Teotitlán Zapotec; tone alternations are also independent of phonation in the language (Uchihara and Gutiérrez 2020). For some verb stems with a glottalized vowel bearing a mid tone, there is an alternation (in personless and first-person verbal forms) between mid tone and falling tone. Example (12) shows that /ḛ/ in the verb stem bears a mid tone when the verb is not marked by person. When the verb has a first-person subject, the /ḛ/ bears a falling tone. The glottalized phonation of /ḛ/ remains unchanged when the tone of /ḛ/ alternates. This indicates that, in Teotitlán Zapotec, tone has a grammatical function independent of phonation.

(12)	No person		1ST person with enclitic		1ST person without enclitic	
	ri-nnḛ̄ːz	"trap"	ri-nnḛ̂ːz=á	"I trap"	ri-nnḛ̂ːz	"I trap"

To summarize, in the majority of Zapotec languages in the survey, phonation types and tones function independently of each other. At the same time, there is great variation in the interaction between phonation types and tones. Some languages have phonation types and tones, which are fully crossed with each other; some have partial overlap in the tonal inventory of non-modal vs. modal phonation types; some have non-modal phonations bearing a subset of the tones borne by the modal phonation; lastly, others have non-modal phonations bearing distinct tone(s) from modal phonation. We also see the role of checked phonation varies across languages: some languages have checked phonation that alternates (either governed by rule or in free variation) with rearticulated phonation; while others have checked phonation in opposition with rearticulated phonation. Generally, in Zapotec, checked is a phonation type, which contrasts with other phonation types independently of tone.

2.4 Phonetics of checked and rearticulated phonation

Sections 2.1 and 2.3 reviewed the phonological properties of checked phonation. The current section goes over the characteristics of the phonetic realization of checked phonation. In order to have a clearer picture of the phonetics of checked phonation, it is useful to compare it to other glottalized phonations, which are phonetically similar. This section thus will also include descriptions of the phonetic properties of rearticulated phonation and creaky/laryngealized phonation (for which the phasing of the glottalization is unspecified) identified in the literature.

2.4.1 F0

As discussed in Section 2.3, the majority of the Zapotec languages contrast phonation types independently of tones. The phonological independence is reflected phonetically in the F0 of the phonation types. The F0 values of the same tone with different phonation types are similar to each other. Crowhurst, Kelly, and Teodocio Olivares (2016) suggests that, in Betaza Zapotec, the two laryngealized phonation types – checked and rearticulated – are realized with similar F0 values when they bear the same tone. Teodocio Olivares (2009) also provided the F0 tracks of different tones with checked, rearticulated, and modal phonations in Betaza Zapotec. The same tones have comparable F0 range and contour in checked (Low and Falling) vs. rearticulated (High, Low, Falling, and Rising) vs. modal (High, Low, Falling, Rising) phonations. Chávez-Peón (2010) compared the F0 of the same tone among different phonations in Quiaviní Zapotec. The comparisons include: High checked vs. High modal; Low & Falling rearticulated vs. Low & Falling modal; High & Low & Falling creaky vs. High & Low & Falling modal. They found that the F0 values (for the modal portion of the vowels) of the same tone are similar between different phonations.

Avelino (2004, 2016) highlighted that for creaky vowels in Yalálag Zapotec, the F0 track has a “trough” pattern. The beginning and the end of creaky vowels have steady F0 track, whereas there is F0 lowering in the middle of the vowels, where glottalization occurs. The F0 values of High and Low tone at the beginning and the end of creaky vowels are similar to those in modal vowels. The two studies by Avelino argue that the beginning and end of the vowels carry the tone for the creaky vowels, making it possible that vowels with non-modal phonation can bear any tones. The argument that the modal portion of vowels contains the tone information whereas the non-modal portion of vowels contains the phonation information has also been proposed by Frazier (2013) for Yucatec Maya. Unlike Yalálag Zapotec, the glottalized phonation in Yucatec Maya does not have contrastive tones and only bears high tone. The tone is expressed at the beginning of the vowel whereas the creaky phonation is realized at the end of the vowel. This finding was in turn confirmed by Avelino (2016).

2.4.2 Voice quality

This section reviews the acoustic descriptions of the voice quality of checked and rearticulated phonation types. The acoustic description by the studies in the survey involves spectral tilt measurements, periodicity measurement, and visual inspection of spectrograms. Spectral tilt measurements, including H1–H2, H1–A1, H1–A3, have been widely used as the acoustic correlates of the degree of vocal fold constriction (Klatt and Klatt 1990; Zhang 2016). Lower spectral tilt values are correlated with a larger degree of vocal fold constriction. A large degree of vocal fold constriction frequently leads to a creaky voice quality. Periodicity of the sound wave can be represented by jitter and shimmer (Brockmann et al. 2011; Heiberger and Horii 1982) and Harmonic-to-Noise Ratio (HNR) (among many others) (Garellek 2019). Higher jitter and shimmer values and lower HNR values indicate the speech signal is noisy. Noise is frequently observed in glottalized and breathy phonations.

In general, when a language contrasts checked phonation with rearticulated phonation, the checked vowels have creakier quality at the end of the vowels, whereas the rearticulated vowels have creakier quality in the middle of the vowels. In Yalálag Zapotec, checked vowels have smaller H1–H2, H1–A1, and H1–A3 values than rearticulated vowels in general. Those values of rearticulated vowels are in turn smaller than those of modal vowels (Avelino 2004). The spectral tilt values were also measured at the 25%, 50%, and 75% of the time-course of the vowel. For H1–A1 and H1–A3, rearticulated vowels reached their lowest point at 50% of the vowels, whereas checked vowels reached their lowest point at 75% of the vowels. The average spectral tilt values and the individual values at different time points of Yalálag vowels indicate that checked and rearticulated vowels were produced with more glottal constriction than modal vowels. Checked and rearticulated vowels differ by the timing of glottalization in the vowels: checked vowels have glottalization at vowel-final positions whereas rearticulated vowels have glottalization in the middle of the vowels.

However, variation in the voice quality of checked and rearticulated vowels is also found in some languages. Teodocio Olivares (2009) and Crowhurst, Kelly, and Teodocio Olivares (2016) provided phonetic descriptions of the checked and the rearticulated vowels in Betaza Zapotec, based on the visual inspection of the spectrograms. In isolation and in phrase-final positions, checked vowels end in a glottal closure (Teodocio Olivares 2009). The vowels sometimes become creaky before the glottal closure (Crowhurst, Kelly, and Teodocio Olivares 2016). Multiple types of creaky voice have been observed in checked vowel realizations[2]. Checked vowels can be realized with vocal fry[3], aperiodicity, or both in the same vowel. In non-phrase-final positions, there is variation in the timing of glottalization in checked vowels. Checked vowels lose their vowel-final glottal closure in non-final positions. Glottalization in checked vowels

2. Crowhurst, Kelly, and Teodocio Olivares (2016) refer to the attributes of creaky voice (e.g. vocal fry and aperiodicity) as "laryngeal markers". For detailed description of the different types of creaky voice, see Keating, Garellek, and Kreiman (2015)

3. Crowhurst, Kelly, and Teodocio Olivares (2016) used the words "creakiness" when describing one type of creak in checked vowels. Based on the spectrograms in Crowhurst, Kelly, and Teodocio Olivares 2016, "creakiness" should probably refer to "vocal fry" per the taxonomy in Keating, Garellek, and Kreiman 2015.

can occur in the middle of the vowels, at the beginning and the end of the vowels, or only at the beginning of the vowels.

Similar to the checked vowels, the realizations of the glottalization in rearticulated vowels in Betaza Zapotec also have variation (Crowhurst, Kelly, and Teodocio Olivares 2016). Typical rearticulated vowels have glottalization in the middle of the vowels. Moreover, there are also tokens where the first third, first half, or first two thirds of the vowels are glottalized respectively, while the rest of the vowel is in modal voice. Both vocal fry and aperiodic voicing have been found in the production of the reariticulated vowels in Betaza Zapotec. Due to the high variability in the timing and type of glottalization in checked vowels and rearticualted vowels, checked and rearticulated vowels can be indistinguishable in non-phrase-final positions.

As described in Section 2.2.3, Quiaviní Zapotec is identified as having “creaky”, “checked”, and “rearticulated” phonations. The “creaky” and “checked” phonations have similar voice quality properties per the description by Chávez-Peón (2010, 2011). Chávez-Peón (2010) stated that “checked” vowels begin with a modal vowel portion, followed by a glottal closure. Based on the visual inspection of the sample spectrograms of the “checked” vowels, Chávez-Peón (2010) found that there is glottalization in the vowels before the glottal closure. After the glottal closure release, there is an optional voiceless vowel that has the same vowel quality as the vowel before the glottal closure, referred to as an “echo vowel” [Vʔ$^{\text{V̥}}$] (Chávez-Peón 2010). Note that the description of “echo vowel” in [Vʔ$^{\text{V̥}}$] is consistent with and aspirated release to the glottal closure (i.e., [V͡ʔ^{h}]). Thus, we can also describe the optional “echo” vowel as an optional aspirated release of the glottal closure.

Chávez-Peón (2010) provided an acoustic analysis of the voice quality of the “creaky” vowels in Quiaviní Zapotec, and compared “creaky” vowels with modal vowels. They measured the H1–H2, H1–A1, and the jitter values of the “creaky” vowels and modal vowels in the language. Low-toned “creaky” vowels have significantly lower H1–H2 and H1–A1 values than modal vowels with all tones in the second half of the vowels, and a higher jitter value on average than modal vowels with all tones. H1–H2 and H1–A1 values are in a falling trend over the timecourse of the “creaky” vowels. The lower H1–H2 and H1–A1 values and the higher jitter values of “creaky” vowels compared with modal vowels suggest that “creaky” vowels have a creakier voice quality than modal vowels. Glottalization occurs at the end of the creaky vowels. Chávez-Peón (2010) also compared the jitter values of “creaky” vowels across different tones. The results showed that, for the female speaker, creaky vowels have lower jitter values in High tone than in Falling tone. This indicates that “creaky” vowels with high tone tend to be less creaky than “creaky” vowels with low and falling tone.

Based on the spectrograms of the “checked” vowels and the acoustic description of the “creaky” vowels in Quiaviní Zapotec, we can see that both “checked” and “creaky” vowels have vowel-final glottalization. They both fit the phonetic definition of a checked phonation. In addition, since no minimal pair has been found between the “checked” and the “creaky” phonations (Chávez-Peón 2010, 2011), I conclude that “checked” and “creaky” phonations might be the variation of the same phonation in Quiaviní, and that phonation can be called “checked.”

As for the rearticulated vowels in Quiaviní Zapotec, Chávez-Peón (2010) suggested that they can be divided into three portions: modal vowel portion, followed by glottal closure, followed by another modal vowel portion, based on the spectrogram inspection. The second modal vowel is obligatory for rearticulated vowels. The glottal closure in rearticulated vowels can be either a full closure or strong glottalization. Since Chávez-Peón (2010) only provided acoustic description of the voice quality for the “creaky” and modal vowels, future studies can measure the voice quality of the “checked” vowels and rearticulated vowels in Quiaviní Zapotec, and test whether “checked” and “creaky” vowels are statistically indistinguishable in production.

Esposito (2003, 2010) concluded that Santa Ana Valle Zapotec has modal, “creaky”, and breathy phonation types. Esposito (2010) compared the modal, “creaky”, and breathy vowels in terms of their H1–H2 and H1–A3 values. H1–A3 distinguished the three phonations for males speakers, whereas H1–H2 distinguished the contrastive phonations for females speakers. There was a hierarchy of breathy > modal > creaky in the spectral tilt value for both groups. Future studies can measure the variation of H1–H2 and H1–A3 at different time points of the vowel, in order to describe the timing of glottalization in “creaky” vowels in Santa Ana del Valle Zapotec.

2.4.3 Duration

Avelino (2004) suggested that in Yalálag Zapotec, the duration of vowels varied according to phonation type as follows: modal vowels closed by voiceless stops and checked vowels in open syllables have the shortest duration (116 ms and 120 ms, respectively), followed by rearticulated vowels (149 ms in closed syllables and 155 ms in open syllables), and modal vowels in open syllables have the longest duration (176 ms).

For Betaza Zapotec (Teodocio Olivares 2009), on the one hand, checked vowels in open syllables have a shorter duration (183 ms) than modal long vowels in open syllables with a lenis onset (201 ms), modal long vowel in closed syllables with a lenis coda (220 ms), and laryngealized vowels in open syllables (240 ms). On the other hand, the duration of checked vowels in open syllables is longer (183 ms) than modal short vowels in open syllables with a fortis onset (139 ms), modal short vowels in closed syllables with a fortis coda (130 ms), and laryngealized vowels in syllables closed by nasals (129 ms).

Chávez-Peón (2010) divided “checked” and rearticulated vowels in Quiaviní Zapotec into three portions – first vowel portion, glottalized portion, and second vowel portion[4] – and measured the duration of each portion. Chávez-Peón (2010) found that, in open syllables, for both the female and the male speakers, the duration of first modal-like vowel portion was comparable between the “checked” and rearticulated vowels. The duration of the second modal-like vowel portion of the “checked” vowels was shorter than the rearticulated vowels. In fact, the second modal vowel portion was frequently omitted in the production, especially when the “checked” vowels were clitics (e.g., /-aˀ/ in /ɾ-ka̰z-aˀ/ “HAB-want-1st” “I want”). The male speaker produced longer glottalized portion for “checked” vowels than for rearticulated vowels. Conse-

4. Chávez-Peón (2010) stated that the echo vowel after the glottal closure in “checked” vowels are usually voiceless. However, per their description of the voicing in the second modal-like vowel, the “echo vowels” of the “checked” vowels still have a certain proportion of voicing, although the proportion of voicing is smaller than the mandatory second modal vowel in the rearticulated vowels.

quently, for the female speaker, the total duration of "checked" vowels (154 ms) was shorter than rearticulated vowels (rearticulated-Low: 192 ms; rearticulated-falling: 180 ms), whereas for the male speaker, "checked" vowels and rearticualted vowels had comparable durations (checked: 184 ms; rearticulated-Low: 179 ms; rearticulated-Falling: 188 ms).

The duration of the "creaky" vowels in Quiaviní Zapotec varies by the syllable type. Before fortis consonants, "creaky" vowels are shorter (127, 109 ms for creaky-H and creaky-L respectively); Before lenis consonants, "creaky" vowels are longer (229, 268 ms for "creaky"-H and "creaky"-L respectively). On average, the total durations of "checked" (in open syllables), rearticulated (in open syllables), and "creaky" vowels (before lenis consonant codas) are comparable to that of the modal vowels (in open syllables or before lenis consonant codas) (averaged duration of modal-High and -Low vowels for the female speaker: 232.5 ms; for the male speaker: 180 ms).

To summarize, Yalálag Zapotec and Betaza Zapotec checked vowels have a shorter duration than modal vowels and rearticulated vowels in open syllables or in environments where the vowels are long, whereas Quiaviní Zapotec "checked" and "creaky" vowels have a similar duration as modal and rearticulated vowels. Such a difference between different languages reflects the realization of the checked phonation is language-specific. The differences in the reported durations can also be attributed to the differences in the segmentation criteria used by different studies. An important decision is whether to count the glottal closure at the end of the checked vowels into the total duration of the checked vowels. Avelino (2004) did not count the glottal closure into the checked vowels, whereas Chávez-Peón (2010) did count the glottal closure and the glottalization as part of the "checked" and the "creaky" vowels. Teodocio Olivares (2009) did not make it explicit whether the glottal closure was counted. However, they differentiated between Vʔ and $Vʔ^h$, the former of which is a checked vowel, while the latter of which is a vowel-plus-glottal stop coda sequence. "Checked" vowels are defined as vowels followed by a complete glottal closure without an aspirated release of the closure. In addition, in the spectrogram of a checked vowel token provided in the paper, the checked vowel was highlighted only for its voiced portion before the glottal closure (Teodocio Olivares 2009, p. 115, Figure 25). Thus, I infer that Teodocio Olivares (2009) did not include the duration of glottal closure into the duration of checked vowels because if there is not a visible release of the glottal closure for checked vowels, the duration of the glottal closure is not measurable. Taking the segmentation criteria of glottal closure into account, we can infer that studies that excluded glottal closure from checked vowels found checked vowels had a shorter duration than modal and rearticulated vowels, whereas studies that included glottal closure into checked vowels found that checked vowels had comparable duration as the modal and rearticulated vowels. In other words, the voiced portion of checked vowels tend to be shorter than the voiced portion of modal and rearticulated vowels (in open syllables).

2.4.4 Vowel quality

The only acoustic study in the survey that compared the vowel quality between modal vowels and rearticulated vowels is Avelino 2004 for Yalálag Zapotec. Avelino (2004) found that there are five modal vowels: /i, e, a, o, u/ but four rearticulated vowels: /iʔi, eʔe, aʔa, oʔo/. The F1 and F2 of /i, e, a, o/ are similar between modal vowels and rearticulated vowels. Modal vowels have lower F1 than rearticulated vowels. For females, modal vowels have more centralized F2 values than rearticulated vowels. However, those differences were not significant.

2.5 Summary of the survey on Zapotec languages

To summarize the survey on Zapotec in terms of checkedness, I found that, the glottal stop in Vʔ and VʔV tend to be suprasegmental. When the glottalization is realized at the late-phase of vowels ([$V^ʔ$]), the phonation can be called “checked” phonation. When the glottalization is realized at the mid-phase of vowels ([$V^ʔV$]), the phonation can be called “rearticulated” phonation. In-phase glottalization is usually a free variation of the rearticulated phonation.

The checked phonation can be either contrastive or allophonic in the language. Both cases have been found in my survey. When checked phonation and rearticulated phonation are in free variation or in complementary distribution of each other, the checked phonation is allophonic. When changing the phasing of glottalization from late-phased to middle-phased changes the word meaning, the checked phonation is contrastive in the language.

In terms of the interaction between checked phonation and tone, I propose six possible interaction types between them, in terms of whether the language contrasts checked phonation with other phonations independently of tone, and whether the language contrasts a tone pair independently of phonation. Type I is when the tone and phonation types fully-crossed with each other. Type II is when the tones of the checked phonation overlap with the tones of other phonations. Type III is when the tone of the checked phontion is a subset of that of other phonations. Type IV is when non-modal phonations bear the same tone, while that tone cannot be borne by the modal phonation. Type V is when the checked phonation can bear multiple tones, but that multiple tones do not overlap with other phonations. Type VI is when the checked phonation has a single designated tone, which is distinct from other phonations. The majority of the Zapotecan languages contrast the checked phonation (along with other non-modal phonations) independently of tone. Phonation can also bear grammatical functions.

The phonological definition of "checked" and its tonally independent nature is also reflected in its phonetic properties. In terms of F0, checked, rearticulated, and modal vowels have similar F0 contours when they bear the same tone. In terms of voice quality, both checked and rearticulated vowels are produced with more glottal constriction than modal vowels. Checked and rearticulated vowels differ in when glottalization occurs in the vowels. Checked vowels have more glottal constriction than rearticulated vowels at the end of the vowels, whereas rearticulated vowels are produced with more glottal constriction in the middle of the vowels. In terms of duration, checked vowels tend to be shorter than rearticulated vowels and modal vowels when they are all in open syllables. In terms of vowel quality, no difference in the vowel quality has been found between modal and checked or rearticulated vowels.

Chapter 3

Survey of checkedness in Chinese languages

Chinese is another language family that has widely-reported "checked" constituents. Section 1.2 provided examples of checked syllables and checked tones in Meixian Hakka Chinese. Meixian Hakka has open syllables and syllables closed by nasals or /-p, -t, -k/[1]. Syllables ending in /-p, -t, -k/ are called "checked" syllables. Checked syllables bear either Tone 55 or 31 (in Chao numerals). Unchecked syllables bear either Tone 31, 33, 51, or 11. The tones that are borne by checked syllables can be called "checked tones." Such checked syllables and tones are frequently observed in Chinese languages. This dissertation defines this type of checkedness as: the phenomenon in which obstruent-closed syllables are associated with distinctive tones, or only a subset of tones, from other syllables. The obstruent-closed syllables are called "checked syllables". The tones borne by such syllables are called "checked tones." The differences in checked constituent between Chinese and Zapotec languages are twofold. They differ in the phonological structure and in their interaction with tone (see Table 3.1). "Checked" in Zapotec languages is a type of phonation, whereas in Chinese languages, the same term refers to a type of syllable. Checked phonation usually contrasts with modal phonation independently of tone in Zapotec languages, whereas in Chinese, checked syllables covary with tones. Tones and checked/unchecked syllables are never fully-crossed with each other. Despite the phonological

1. In this section, I use hyphen plus consonant (e.g., -p) to indicate the consonant is in the coda position of a syllable, and use consonant plus hyphen (e.g., p-) to indicate that the consonant is in the onset position of a syllable.

differences between checked phonation and checked syllables/tones, the label of "checked" is assigned to both cases because they both have the phonetic characteristics of "being checked." Checked phonation has an abrupt offset of voicing by the glottalization at the end of vowels. Checked syllables/tones also have the voicing of the syllable nuclei shut abruptly by the voiceless stop coda. It is the phonetic nature of "being checked" that units those two distinct phonologically constituents.

Table 3.1. Comparison between checked syllables in Chinese languages and checked phonation in Zapotec languages

Zapotec	Chinese
Checked phonation /Vˀ/	Checked syllable/tone /Vp, -t, -k, -ʔ/
Tonally-independent	Tonally-dependent

Let's now revisit the question proposed at the beginning of this dissertation: is checked syllable and checked tone a phonological constituent that is distinct from short duration, glottalized phonation, closed syllable, and tone? I maintain that "checked" is a distinct phonological constituent. It is a descriptor of the phonotactic constraint between closed syllables and tones. The descriptor of "checked" is necessary because neither the category of closed syllable nor the category of tone alone can explain the phonotactic constraint between closed syllable and tones. First, The concept of "closed syllable" does not explain why closed syllables in languages like Meixian Hakka differ from nasal-closed syllables by bearing different set of tone. Second, The concept of "tone" does not explain why for languages like Meixian Hakka, certain tone can only be associated with closed syllables. To summarize, "checked" in Chinese languages is neither a phonological feature, nor a phonological category, but a label for a phonotatic restriction. The concept "checked" can be reduced to the combination of two phonological categories – closed syllable and tone.

This section illustrates the checkedness in Chinese languages from three perspectives.

First, I review the origin of checkedness in Chinese – **Rù* syllables[2] and **Rù* tones in Middle Chinese. Second, I compare checked syllables and tones with **Rù* syllables and tones, and distinguish those two sets of concepts. Last, I review the phonetic properties of checked syllables and tones in Chinese languages.

3.1 Tonogenesis from Old Chinese to Modern Chinese

The term "checked" is closely related to the term "entering" or *"*Rù"* 入 in Chinese linguistic tradition. Many studies thus use "checked" interchangeably with "entering" and *"*Rù"* (Sung 1973; Xu 2020; Zhang 2017; Zhang 2006; Pan 2017; Chan 1987). In fact, based on the definition of checkedness given in this dissertation, the majority of checked syllables and tones today originate from the **Rù* syllables and **Rù* tones in Middle Chinese.

However, this dissertation will not equate the term "checked" with *"*Rù"*. "Checked" refers to a synchronic phenomenon, whereas *"*Rù"* refers to a phonological category in Middle Chinese. A detailed comparison and contrast between checked and **Rù* will be given in Section 3.4. In the current section, I will review the diachronic origin of **Rù* syllables 入声韵 and **Rù* tones 入声调 and the reflex of **Rù* syllable and **Rù* tone in Modern Chinese. The definitions of **Rù* and tone, and checked syllable and tone are as follows:

- **Rù* syllable: Syllables that are reconstructed as closed by /*-p, *-t, *-k/ in Middle Chinese;
- **Rù* tone: Tones that are reconstructed as accompanying **Rù* syllables in Middle Chinese;
- Checked syllable: Syllables that are closed by obstruents and associated with distinctive tones from open syllables and sonorant-closed syllables, or a subset of tones borne by open syllables and sonorant-closed syllables;
- Checked tone: Tones that are borne by checked syllables.

2. The asterisk "*" indicates that the form is constructed.

Chinese language is periodized into four periods: Old Chinese 上古汉语 (12th century BC – 3rd century AD); Middle Chinese 中古汉语 (4th – 12th centuries AD); Early Modern Chinese 近代汉语 (13th – early 20 centuries AD), and Modern Chinese 现代汉语 (Early 20th centuries – now) (Dong 2020, p. 7). Three of those periods – Old Chinese, Middle Chinese and Modern Chinese – are of particular interest for the discussion of **Rù* syllables and tones and checked syllables and tones in this dissertation.

As first proposed by Haudricourt (1954), Proto-Vietnamese, Tai, and Old Chinese are toneless. Since Old Chinese, Chinese languages have gone through two stages of tonogenesis. The first stage is conditioned by the loss of the syllable coda /*-ʔ/ and /*-h/ in Middle Chinese. Three phonemic tones were generated in Middle Chinese: **Píng* 平, **Shǎng* 上, and **Qù* 去 tones, corresponding to Old Chinese syllables ending in /*-∅, *-N/, /-ʔ/, /-h/, respectively. In Middle Chinese, there is another tone category – **Rù* 入, which are associated with syllables ending in /-p, -t, -k/. However, The **Rù* "tone" is actually considered toneless, because it is already distinguished from other syllables and tones by its syllable structure (Haudricourt 1954; Ratliff 2015).

Píng* and **Shǎng* tones are reconstructed as (low) level and rising tones respectively. There are debates of what **Qù* refers to: Ferlus (2009) argues that **Qù* tone refers to a breathy voice quality in the vowel due to the reminiscent of the /*-h/ coda in the syllables that were lost in Old Chinese. Sagart (1986) maintained that **Qù* tone is glottalized at some stage in the history and creates a falling pitch percept due to the glottalization in the middle of the vowel. Chen (1976) denoted that Qù tone has a falling pitch. Lastly, **Rù* tone does not refer to melody, but to a syllable structure. It refers to syllables that end in obstruents /p, t, k/. Such kind of syllables do not have contrastive tones, and have characteristics of being "short" and "abrupt" (Sagart 1993; Chen 1976; Norman 1988a; Mei 1970). The phonetic nature of the four tones in Middle Chinese is backed up by the writings from scholars and monks in the history (Mei 1970). There are two quotes from 9th century and Ming dynasty, as cited from Mei (1970): "平声哀而安，上声厉而举，去声清而远，入声直而簇。(Píng* is sad and peaceful. **Shǎng* is harsh and rising. *Qù* is clear and far. *Rù* is straight and abrupt)"; "平声平道莫低昂，上声高呼猛烈强，去声分明哀远道，入声短促急收藏。(*Pìng* is level and non-low. *Shǎng* is produced high and intensely. *Qù* is clear and far away. *Rù* is short, abrupt, and ended quickly.)"

The second stage of tonogenesis is tonal split conditioned by the onset of syllables. At certain stage of Middle Chinese, the voicing contrast in the onset was lost. The four tones (including the *Rù* tone) are split into eight tones by the voicing of the onset. Syllables with a voiceless onset acquire the upper register – *Yīn* 阴, wheres syllables with a voiced onset acquire the lower register – *Yáng* 阳. The two stages of tonogensis is schematized in Table 3.2. It is adapted from Table 16.1 from Ratliff 2015 and the table from Zhengzhang 2003, p. 219.

Table 3.2. Tonogenesis from Old Chinese to Middle Chinese

Stage 1 Old Chinese	/*-∅, -l, -m, -n, -ŋ/	/*-ʔ/	/*-s/ > /*-h/	/*-b, -d, -g/
Stage 2 Middle Chinese (tonogenesis)	*Píng* (level)	*Shǎng* (rising)	*Qù* (falling)	*Rù* (atonal) /*-p, t, k/
Stage 3 Middle Chinese (tone split)	*t-: *Yīnpíng* *d-: *Yángpíng*	*t-: *Yīnshǎng* *d-: *Yángshǎng*	*t-: *Yīnqù* *d-: *Yángqù*	*t-: *Yīnrù* *d-: *Yángrù*

Ferlus (2009) and Ratliff (2015) stated that the *Rù* tone became tonal when its carrier syllable lost the /-p, -t, -k/ coda. An example is White Hmong creaky tone, which is associated with /*-p, *-t, *-k/-closed syllables in Proto-Hmong, and now is realized with creaky voiced Tone [21] after the /*-p, *-t, *-k/ codas are lost. However, in Modern Chinese, there are many varieties that preserve the /*-p, *-t, *-k/ codas (or a subset of them). At the same time, the /-p, -t, -k/-closed syllables are consistently produced with certain pitch contours. Strictly speaking, the pitch contours associated with the /-p, -t, -k/-closed syllables do not contrast with the tones associated with other syllables independently. However, studies on Chinese still identifies the pitch contours associated with the /p, t, k/-closed syllables as "tones." One possible reason is the *Yīn* and *Yáng* register split in the /*-p, *-t, *-k/-closed syllables in the later stage of Middle Chinese. After the second stage of tonogenesis, for the /*-p, *-t, *-k/-closed syllables, the *Yīng* and *Yáng* tonal registers became contrastive, making the **Rù* category to be tonal consequently. Despite the ambiguity of whether **Rù* tone is tonal, and at what stage **Rù* tone can be counted as being tonal, literature of Chinese linguistics distinguishes the concept of tone and syllable structure for the **Rù* category. In Middle Chinese, syllables closed by /*-p, *-t, *-k/ are called **Rù* syllables 入声韵. The pitch contour(s) associated with the **Rù* syllables are called **Rù* tone 入声调.

To illustrate the path of tonogenesis from Old Chinese to Modern Chinese, I present sample words in three varieties of Modern Chinese (Heyuan 河源 Hakka, Chaling 茶陵 Gan, Beijing Mandarin), and their reconstructions in Old Chinese and Middle Chinese. For the words of the Modern Chinese varieties, I mark their reconstruction of the eight tones in Middle Chinese after the second stage of tonogenesis, as provided in the references. The reconstructions of all the examples in Old Chinese and Middle Chinese provided in this dissertation are from Zhengzhang 2003. The examples show that the eight tones in Middle Chinese do not have one-on-one correspondence in Modern Chinese. For example, **Yángshǎng* is merged into **Yīngqù* and **Yángqù* in Heyuan Hakka and Chaling Gan, respectively. **Yīnrù* and **Yángrù* are merged into one tone in Chaling Gan. **Yángshǎng*, **Yīnqù* and **Yángqù* are merged as **Qù* tones in Beijing Mandarin. Heyuan Hakka, Chaling Gan, and Beijing Mandarin reflect different stages of the sound change of **Rù* syllables and **Rù* tones as well. The /*-p, *-t, *-k/ codas in **Rù* syllables are preserved in Heyuan Hakka, but lost in Chaling Gan and Beijing Mandarin. The **Rù* tones are preserved in Heyuan Hakka and Chaling Gan, but are lost in Beijing Mandarin.

Table 3.3. Examples of the tones in Heyuan Hakka, Chaling Gan, Beijing Chinese, and their reconstruction in Old Chinese and Middle Chinese. The examples for Modern Heyuan Hakka and Chaling Gan are from Li et al. 1992. Their reconstructions in Old Chinese and Middle Chinese are from Zhengzhang 2003.

	餐 "meal"	残 "incomplete"	朵 "flower"	柱 "pillor"	过 "pass"	大 "big"	答 "answer"	及 "and"
Old Chinese	*/shaan/	*/zlaan/	*/toolʔ/	*/doʔ/	*/klools/	*/daads/	*/tkuub/	*/grɯb/
Middle Chinese	*/tsʰan/ *Píng*	*/dzan/ *Píng*	*/tua/ *Shǎng*	*/ɖɨo/ *Shǎng*	*/kua/ *Qù*	*/kai/ *Qù*	*/tʌp/ *Rù*	*/gɣiɪp/ *Rù*
Heyuan Hakka	/tsʰan 33/ **Yīnpíng*	/tsʰan 31/ **Yángpíng*	/tuɔ 24/ **YīnShǎng*	/tsʰy 12/ **Yīnqù*	/kuɔ 12/ **Yīnqù*	/tʰai 55/ **Yángqù*	/tap 5/ **Yīnrù*	/kʰip 2/ **Yángrù*
Chaling Gan	/tsʰã 353/ **Yīnpíng*	/tsʰã 424/ **Yángpíng*	/to 53/ **YīnShǎng*	/tsʰy 12/ **Yángqù*	/kuo 33/ **Yīnqù*	/tʰæ 12/ **Yángqù*	/ta 21/ **Rù*	/tɕʰie 21/ **Rù*
Beijing Mandarin	/tsʰan 55/ **Yīnpíng*	/tsʰan 35/ **Yángpíng*	/two 213/ **Shǎng*	/ʈʂu 51/ **Qù*	/kwo 51/ **Qù*	/ta 51/ **Qù*	/ta 35/ **Yángpíng*	/tɕi 35/ **Yángpíng*

The examples in Table 3.3 reflect that there is variation in the reflexes of the Middle Chinese syllables and tones in Modern Chinese. The research question of the current section is: how are the Middle Chinese **Rù* syllable and **Rù* tone reflected in Modern Chinese, and how are they related to the checked syllables and checked tones defined in this dissertation? To thoroughly answer the question above, I survey the syllable structure and tone system of 95 Chinese language varieties from 57 studies. The list of the languages and references are in Table 3.4. The survey covers the Chinese language families of Yue 粤, Wu 吴, Gan 赣, Min 闽, Xiang 湘, Jianghuai Mandarin 江淮官话, and Southwest Mandarin 西南官话. Within each language, I investigate the present-day reflexes of **Rù* syllables and **Rù* tones. Following the convention in Chinese literature, tones are transcribed using Chao numerals (Chao 1930) unless otherwise specified. Numeral 1 represent the lowest pitch and numeral 5 represents the highest pitch level (within normal range). Most of the words are represented by two Chao numerals, the first of which indicates the pitch of the rhyme onset; the second of which indicates the pitch of the rhyme offset. Some studies assign only one Chao numeral for tones that are short.

Table 3.4. Languages in the survey of Chinese languages.

Family	Language	Source
Gan	Chongyang 崇阳	Zhu 2015
	Chaling 茶陵, Yongxin 永新, Jishui Luotian 吉水螺田 Liling Baitutan 醴陵白兔潭, Xinyu Shatu 新余沙土 Shaowu 邵武, Pingjiang Nanjiang 平江南江, Xiushui 修水 Anyi 安义, Duchang 都昌, Yangxin Guohe 阳新国和 Susong Heta 宿松河塔, Yugan 余干, Yiyang 戈阳 Nancheng 南城, Jianning 建宁	Li et al. 1992
	Wuning 武宁	Sagart 1990
	Yueyang Xinqiang 岳阳新墙, Liuyang Jiaoxi 浏阳焦溪 Pingjiang Meixian 平江梅仙, Liling Lujiang 醴陵渌江	Li 2005
	Pingjiang Chengguan 平江城关	Zhang 2014
	Xupu Longtan 溆浦龙潭	Li 2007
Hakka	Meixian 梅县, Wengyuan 翁源, Liannan 连南 Heyuan 河源, Qingxi 清溪, Jiexi 揭西, Xiuzhuan 秀篆 Wuping Yanqian 武平岩前, Changting 长汀, Ninghua 宁化 Ningdu 宁都, Sandu 三都, Ganxian Panlong 赣县蟠龙 Dayu 大余, Xihe 西河, Luchuan 陆川 Hong Kong Xigong Hakka 香港西贡客家话	Li et al. 1992
	Luoyuan She 罗源畲族	Norman 1988b
	Xin'an 新安	Chappell and Lamarre 2005
	Meixian 梅县	Shao 2012 Lee and Zee 2009
Jianghuai Mandarin	Nanjing 南京	Chen and Wiltshire 2013 Yang and Chen 2018 Oakden 2017
	Anqing 安庆, East Hefei 肥东, Wuhu 芜湖	Tang 2014
	Nantong 南通	Ao 1993
	Ningbo 宁波	Qian 1990
	Rugao 如皋	Xu 2020
Southwest Mandarin	Kunming 昆明	Pinson and Pinson 2000
	Lizhuang 李庄	Yang 1987
Jin	Huhetaote 呼和浩特, Fengzhen 丰镇, Baotou 包头 Shangdu 商都, Baochang 宝昌, Dongsheng 东胜	Liu 2012
	South Taiyuan 太原南郊	Jia 2013
Min	Jianyang 建阳, Jianou 建瓯, Shaowu 邵武 Pucheng 浦城, Ding'an 定安	Norman 1969
	Shaxia Gaizhu 沙县盖竹	Deng 2007

continued …

Table 3.4. Languages in the survey of Chinese languages (continued.)

Family	Language	Source
	Pingxi 平溪	Hu 2020
	Hainan Wenchang 海南文昌	Woon 1979a, 1979b
	Amoy 厦门	Lai 2016 Sung 1973
	Danzhou Hainan 海南儋州	Ting 1980
	Taiwanese 台湾	Kuo 2013 Pan 2017
	Northern Taiwan 台湾北部	Tung, Chao, and Lan 1967
	Zhangzhou 漳州	Huang 2018
	Yun'ao 云澳	Zhang 2017
	Putian Daitou 莆田埭头	Chen 2017
	Putian Jiangkou 莆田江口	Huang 2013
	Gutian Pinghu 古田平湖	Chen 2006
	Fuzhou 福州	Chan 1993, 1997 Shao 2012
Wu	Shanghainese 上海	Zee and Xu 2017 Chen and Gussenhoven 2015
	Shaoxing 绍兴	Zhang 2006
	Luqiao 路桥	Lin 2011
	Ningbo 宁波	Lyu 2019
Xiang	Changsha 长沙	Shao 2012 Li and Liu 2006
	Chishan 赤山	Liu 2013
Yue	Cantonese	Bauer and Matthews 2017 Cheung 1986 Chan 1987 Qin and Mok 2014 Zhu et al. 2008
	Taishanese 台山	Cheng 1973 Tan 2016
	Shijie 石碣, Wangniudun 望牛墩, Dongkeng 东坑 Humenchigang 虎门赤岗, Hengli 横沥 Zhongtangjiangnan 中堂江南, Tangxia 塘厦 Gaobuxiansha 高埗冼沙, Wanjiang 万江	Li 2010

3.2 Reflexes of **Rù* syllables in Modern Chinese

Middle Chinese **Rù* rhymes by definition are the rhymes that are closed by a voiceless obstruent /*-p/, /*-t/, or /*-k/. Compared to Middle Chinese, contemporary Chinese languages show augmentation of the contrast to /p, t, k, ʔ/, full preservation of the contrast, or – most commonly – simplifcation of the set of contrasts. Simplification of the contrasts includes both historical mergers (e.g. /*-p/ and /*-t/ > /-t/), debuccalization (/*-p, *-t, *-k/ > /-ʔ/), and coda loss (/*-p, *-t, *-k/ > ∅). The survey found 15 obstruent coda configurations in Modern Chinese as the reflexes of Middle Chinese **Rù* syllables. The survey results are summarized in Table 3.5. Note that even if a language variety has obstruent codas as the reflexes of Middle Chinese /*-p, *-t, *-k/, it does not mean that every case of Middle Chinese /*-p, *-t, *-k/ still correspond to an obstruent-closed word in that Modern Chinese variety. Thus where Table 3.5 shows that the reflex of Middle Chinese /*-p, *-t, *-k/ in a Modern Chinese variety is /-p, -k/, it means that there exist cases where **Rù* syllables in Modern Chinese become syllables ending in /-p, -k/ in a present-day variety. For other words in the language, it is possible that reflexes of the Middle Chinese /*-p, *-t, *-k/ codas are sonorants, or have been deleted. Only if a present-day language variety does not have any obstruent coda, do I list the obstruent reflex of **Rù* syllable as "∅" in that language. Also note that these 15 patterns are the ones found for the 95 languages in the survey; however, it is quite possible that other patterns can be found upon surveying additional languages.

Table 3.5. Reflexes of Middle Chinese **Rù* syllables in Modern Chinese.

Obstruent group	Dialect group	Example	Checked syllable?
p, t, k	Jianning Gan 建宁	/tap 2/<*/tʌp/ 答 "answer" /tiet 2/<*/det/ 跌 "fall down" /pɔk 2/<*/pɑk/ 博 "broad"	Yes
p, t	Xiuzhuan Gan 秀篆	/tap 24/<*/tʌp/ 答 "answer" /tɛt 24/<*/det/ 跌 "fall down" /pɔu 24/<*/pɑk/ 博 "broad"	Yes

continued …

Table 3.5. Reflexes of Middle Chinese **Rù* syllables in Modern Chinese (continued.)

Obstruent groups	Dialect groups	Examples	Checked syllable?
p, k	Zhongtang Jiangnan Yue 中堂江南	/lɐᵘp 3/<*/nʌp/ 纳 "accept" /tʃik 3/<*/tset/ 节 "festival" /pak 3/<*/bɣæk/ 白 "white"	Yes
t, k	Wengyuan Hakka 翁源	/tak 2/<*/tʌp/ 答 "answer" /tiet 2/<*/det/ 跌 "fall down" /pɔk 2/<*/pɑk/ 博 "broad"	Yes
k	Wangniudun Yue 望牛墩	/ŋak 3/<*/lɑp/ 腊 "December" /ŋᴇk 3/<*/lɑt/ 辣 "spicy" /mok 3/<*/muk/ 木 "wood"	Yes
tn, kŋ	Yugan Gan 余干	/tatn 21/<*/tʌp/ 答 "answer" /tietn 21/<*/det/ 跌 "fall down" /pɔkŋ 21/<*/pɑk/ 博 "broad"	Yes
p, t, k, ʔ	Hainan Wenchang Min 海南文昌	/tiap 42/<*/dʑiɛp/ 涉 "involve" /tiat 42/<*/ʑiɪt/ 实 "concrete" /tiak 42/<*/dʑɨuk/ 熟 "cooked" /beʔ 42/<*/bɣæk/ 白 "white"	Yes
p, t, ʔ	Anyi Gan 安义	/tɔp 5/<*/tʌp/ 答 "answer" /tʰat 2/<*/dɑt/ 达 "arrive" /pɔʔ 53/<*/pɑk/ 博 "broad"	Yes
p, k, ʔ	Yun'ao Min 云澳	/tsap 43/<*/tɕiɪp/ 汁 "juice" /soʔ 43/<*/siuᴇt/ 雪 "snow" /tik 43/<*/tək/ 得 "get" /tiʔ43/<*/tek/ 滴 "drop"	Yes
t, k ʔ	Sandu Hakka 三都	/tat 1/<*/tʌp/ 答 "answer" /tiɛt 1/<*/det/ 跌 "fall down" /pʰɔk 1/<*/pɑk/ 博 "broad" /luʔ 5/<*/ȵɨok/ 辱 "humiliate"	Yes
t, ʔ	Jishui (Luotian) Gan 吉水 (螺田)	/tat 2/<*/tʌp/ 答 "answer" /tiɛʔ 2/<*/det/ 跌 "fall down" /poʔ 2/<*/pɑk/ 博 "broad"	Yes
k, ʔ	Dongkeng Yue 东坑	/nak 3/<*/nʌp/ 纳 "accept" /tʃeʔ 23/<*/tset/ 节 "festival" /pak 3/<*/bɣæk/ 白 "white"	Yes
k, l	Duchang Gan 都昌	/tal 45/<*/tʌp/ 答 "answer" /tiɛl 45/<*/det/ 跌 "fall down" /pɔk 45/<*/pɑk/ 博 "broad"	Yes
ʔ	Wuning Gan 武宁	/lœʔ 42/<*/nʌp/ 纳 "accept" /pæʔ 42/<*/pɣɛt/ 八 "eight" /pɑʔ 223/<*/bɣæk/ 白 "white"	Yes
∅	Lizhuang Southwest Mandarin 李庄西南官话	/ta 24/<*/tʌp/ 答 "answer" /pa 24/<*/pɣɛt/ 八 "eight" /tso 24/<*/ʈɣʌk/ 桌 "desk"	No

Despite the great variation in the precise reflexes of **Rù* across contemporary Chinese, the general trend is that the historical set of **Rù* codas is reduced or lost. The different coda patterns in the reflexes of **Rù* syllables indicates that different languages are at different stages of the obstruent coda loss. Based on the degree of Middle Chinese **Rù* syllable coda loss, I categorize the fifteen coda patterns in Table 3.5 into six types in Table 3.6.

Table 3.6. Categorization of Middle Chinese checked rhymes in contemporary Chinese

Type	Coda	Example
Type I: Full preservation	/-p, -t, -k/	Jianning Gan
Type II: Partial preservation	/-p, -t/	Xiuzhuan Gan
	/-p, -k/	Zhongtang Jiangnan Yue
	/-t, -k/	Wengyuan Hakka
	/-k/	Wangniudun Yue
	/-tn, -kŋ/	Yugan Gan
Type III: Partial debuccalization	/-p, -t, -k, -ʔ/	Hainan Wenchang Min
	/-p, -t, -ʔ/	Anyi Gan
	/-p, -k, -ʔ/	Yun’ao Min
	/-t, -k, -ʔ/	Sandu Hakka
	/-t, -ʔ/	Jishui (Luotian) Gan
	/-k, -ʔ/	Dongkeng Yue
	/-k, -l/	Duchang Gan
Type IV: Full debuccalization	/-ʔ/	Wuning Gan
Type V: Complete loss	∅	Lizhuang Southwest Mandarin

As shown in Table 3.5, there are languages that preserve all three /*-p, *-t, *-k/ codas in Middle Chinese **Rù* syllable, such as Cantonese and Meixian Hakka. I call those languages as Type I – complete preservation of the obstruent codas in Middle Chinese.

Type II languages are those that have a reduced number of /*-p, *-t, *-k/ contrast in their reflex of **Rù* rhymes, including /p, t/, /p, k/, /t, k/, and /k/. The reduction in coda contrast is usually a result of neutralization among /*-p, *-t, *-k/. The general trend of the coda neutralization is for /*-p/ and /*-t/ to merge with /*-k/, yielding only /-k/ (Li 2010). For example, as Table 3.5 shows, Wengyuan Hakka has /*-p/ becoming /-k/; Zhongtang Jiangnan Yue has /*-t/ becoming /-k/; and Wangniudun Yue has /*-p, *-t/ becoming /-k/. One exception is Xiuzhuan Gan, in which /*-p/ and /*-t/ are preserved, but /*-k/ is deleted. Type II has an outlier – Yugan Gan. The reflex of **Rù* syllable coda /p, t, k/ in Modern Yugan Gan are nasally-released /tn/ and /kŋ/. It is unclear what those codas derive from diachonically. According to Li et al. (1992), the nasals are only produced when the words are at phrase-final position. Although /tn/ and /kŋ/ are not obstruents in the narrow sense, I included /tn/ and /kŋ/ in the obstruent reflex of **Rù* syllable because they do bear tones that are different from syllables that are open or closed by other codas. Such kind of syllable structure and tone configuration greatly resemble the **Rù* syllable and **Rù* tone in Middle Chinese.

Type III languages went through partial debuccalization of the oral codas /*-p, *-t, *-k/. There are configurations of /p, t, k, ʔ/, /p, t, ʔ/, /p, k, ʔ/, /t, k, ʔ/, /t, ʔ/, and /k, ʔ/. As shown in Table 3.5, Hainan Wenchang Min has some /*-k/ debuccalized into /ʔ/, but other /*-k/ preserved as /k/. Anyi Gan, and Sandu Hakka has /*-k/ becoming /-ʔ/; Yun'ao Min and Jishui (Luotian) Gan has /*-t/ and /*-k/ becoming /-ʔ/; Dongkeng Yue has /*-t/ becoming /-ʔ/; Wuning Gan has /*-p, *-t, *-k/ all becoming /-ʔ/. It is very likely that the lenition happens in an order of /*-p/ and /*-t/ merging into /*-k/ first, then /*-k/ is lenited to /*-ʔ/. For example, in Dongkeng Yue, we see examples of /*-p/ becoming /-k/ (/nak 3/<*/nʌp/); /*-t/ becoming /-ʔ/ (/tʃeʔ 23/<*/tset/), and /*-k/ remaining to be /-k/ (/pak 3/<*/bɣæk/). However, this hypothesis has to be confirmed by the data of the languages from different periods in the history.

The sound change from /*-p, *-t, *-k/ to /-ʔ/ reflects the debuccalization of the **Rù* syllable obstruent coda. There are two possible paths for the debuccalization. The first hypothesis is that /*-p, *-t, *-k/ > /-ʔ/ directly. This implies that codas lose their oral place but retain their [+consonantal] feature. The second hypothesis is that /*-p, *-t, *-k/ were historically glottalized (i.e., [*-ˀp], *-ˀt], *-ˀk]). Then, the oral place is lost, but the glottalization is preserved in the reflexes of **Rù* syllables.

There is an outlier language in Type III languages – Duchang Gan. Although /l/ is a sonorant, I included /l/ in the obstruent reflex of Middle Chinese /p, t, k/ coda in Duchang Gan (Li et al. 1992). The reasons are that in Duchang Gan, 1) /l/ are derived from words ending in obstruents in Middle Chinese, and 2) words closed by /l/ carry the same set of tone as words closed by /k/. Examples in Table 3.5 show that /-l/ in Duchang Gan is derived from /*-p/ and /*-t/ (/tal 45/<*/tʌp/; /tiɛl 45/<*/det/).

Type IV coda configuration demonstrates complete debuccalization in the reflexes of the obstruent coda in Middle Chinese. The only licit obstruent coda in languages of Type IV is the glottal stop. In those languages, Middle Chinese /*-p, *-t, *-k/ are lenited to the glottal stop. Examples have been found from Gan, Hakka, Min, Wu, Jin, and Jianghuai Mandarin.

Type V languages have no obstruent coda allowed in the languages. All Middle Chinese *Rù* rhymes have lost their [+obstruent] feature in the coda completely. To illustrate the relation between Middle Chinese **Rù* syllable and their reflexes in Modern Chinese more directly, I selected four words that have **Rù* syllables in Middle Chinese and listed their transcription in Middle Chinese and in eight Modern Chinese varieties in Table 3.7.

Table 3.7. Reflexes of Middle Chinese **Rù* syllables in Modern Chinese in four words in eight Modern Chinese varieties. The segments in parenthesis besides the language name indicate the licit obstruent codas in the language. Data for Tangxia Yue are from Li 2010. All other data are from Li et al. 1992.

	十 "ten"	末 "end"	八 "eight"	落 "fall"	Checked syllable?
Middle Chinese	/*dʑiɪp/	/*muɑt/	/*pɣɛt/	/*lɑk/	Yes
Meixian 梅县 Hakka (-p -t -k)	/səp 5/	/mat 5/	/pat 1/	/lɔk 5/	Yes
Tangxia 塘厦 Yue (-p -t -k -ʔ)	/ɬɐp 3/	/mut 3/	/piæᵃʔ 13/	/luɔk 3/	Yes
Wengyuan 翁源 Hakka (-t -k)	/sit 5/	/mat 5/	/pat 2/	/lɔk 5/	Yes
Yugan 余干 Gan (-tn -kŋ)	/ɕiɛtn 1/	/motn 5/	/patn 5/	/lɔkŋ 1/	Yes
Duchang 都昌 Gan (-k -l)	/ʂəl 3/	/mɔl 45/	/bal 45/	/lɔk 3/	Yes
Xiushui 修水 Gan (-t -ʔ)	/sɤt 32/	/mɔt 32/	/pat 32/	/loʔ 32/	Yes
Yiyang 弋阳 Gan (-ʔ)	/sɛʔ 5/	/moʔ 5/	/paʔ 5/	/laʔ 5/	Yes
Changling 茶陵 Gan (∅)	/ʂʅ 21/	/mo 21/	/pa 21/	/lo 21/	No

3.2.1 Status of glottal stop in reflexes of the **Rù* syllable

For languages that have Vʔ-structured syllables, the same question arises as for the Vʔ syllables in Zapotec languages – is the glottal stop as segment or a suprasegment in those languages? In this section I review the reasons proposed in the literature. In general, the arguments are made by comparing the phonological structure of Vʔ with with open syllables (V) and syllables closed by nasals (VN). If Vʔ behaves similarly to open syllables, the glottal stop is more likely to be a suprasegment. If Vʔ behaves similarly to VN, the glottal stop is more likely to be considered a segment. The comparisons that are found in the survey come from two perspectives: 1) comparing the vowel quality between Vʔ and open syllable and closed syllables; 2) comparing the onset assimilation patterns between Vʔ and open syllable and closed syllables.

3.2.1.1 Evidence of glottal stop status from vowel quality

Two languages provide examples for the arguments for and against glottal stop as a segment based on the vowel quality in CV, CVʔ, and CVN. Zhang (2006) proposes that the glottal stop in CVʔ in Shaoxin Wu is a consonantal segment because CVʔ syllables have the same vowel quality as CVN, but different vowel quality from CV. Sagart (1990) proposes that the glottal stop in CVʔ in Wuning Gan is a suprasegment because CVʔ syllables have the same vowel quality as CV, but different vowel quality from CVN and CVV (syllables with diphthong). I argue that the first argument is strong evidence for glottal stop as a segment, whereas the second argument is definitive for glottal stop as a suprasegment.

3.2.1.1.1 Shaoxing Wu: CVʔ pairing with CVN and differing from CV in vowel quality

Zhang (2006) proposes that the glottal stop in Vʔ in Shaoxing Wu is a segment, because syllables closed by either a glottal stop and or a nasal exhibit vowel laxing (becoming more centralized or lower), whereas only tense (more peripheral) vowels are allowed as the nucleus in open syllables. The vowel distribution by syllable structure is shown in Table 3.8. Zhang (2006) argues that this is due to a requirement of syllable weight. Syllables in Shaoxing Wu are bimoraic. Tense vowels are bimoraic; lax vowels and codas are monomoraic; and onset is non-moraic. Consequently, tense vowels can be the nucleus of open syllables, whereas lax vowels do not occur as the nucleus of open syllables, but can only be in closed syllables. Thus, lax vowels being allowed in Vʔ speaks to the glottal stop being a segment in Shaoxing Wu.

Table 3.8. Distribution of vowel by syllable type in Shaoxing Wu (Zhang, 2006, p.162)

V(V)	VN	Vʔ
/i/	/ɪŋ/	/ɪʔ/
/ʅ/		
/y/		
/e/		/ɛʔ/
/a/[a]	/aŋ/	/aʔ/
/ɤ/		
/o/[a]	/oŋ/	/oʔ/
/u/		
/ɑɒ/	/ɒŋ/	
	/əŋ/	/əʔ/

[a] Zhang (2006) stated that /a/ and /o/ are unspecified for the [+tense] feature. Thus, /a/ and /o/ are allowed in all three environments.

3.2.1.1.2 Wuning Gan: CVʔ pairing with CV and differing from CVN in vowel quality

Sagart (1990) also appeals to vowel quality when determining the status of glottal stop in Vʔ syllables in Wuning Gan. In Wuning Gan, /æ/ and /œ/ are centralized to [a] and [ɵ], respectively, when they are part of a diphthong or when they occur before a nasal. In open syllables or before glottal stop, the quality of /æ/ and /œ/ does not change. Sagart (1990) uses this as the evidence for glottal stop being a suprasegment because Vʔ behaves the same as open syllables in terms of this vowel quality alternation. However, this argument, while plausible, is not definitive. It is also possible to treat the glottal stop in Vʔ as a segment by attributing the lack of vowel quality alternation in Vʔ to the "placeless" and [-sonorant] nature of glottal stop. As discussed in Section 2.1.3, glottal stops can be transparent to phonological alternations and behave differently from other consonants due to their lack of supralaryngeal place feature (i.e., being "placeless"). Applying this reasoning to Wuning Gan, we can argue that glottal stop is a segment in Wuning Gan by positing a rule such that vowel centralization occurs only before an oral segment with a specified supralaryngeal place of articulation. When the vowel is followed by a placeless consonant ʔ, the vowel quality remains unchanged, because there is no motivation for a change of place of articulation in Vʔ sequence. Duanmu (1994) used the same argument to explain why, in Garo, vowel alternations (/i/ → [ɪ]; /e/ → [i]; /o/ → [u]) do not apply to open and Vʔ syllables but do apply in syllables closed by consonants other than glottal stop. Another possible reasoning is that the vowel alternation caused by being in a bimoraic syllable. Syllables with VV (diphthong) or VN are treated as bimoraic because the other vowel in the diphthong and the nasal are sonorants. Vʔ is treated as monomoraic because the glottal stop is not sonorous. Using either the reasoning of glottal stop being placeless consonant or the reasoning of glottal stop being not sonorant, we can explain why Vʔ behaves differently from VV and VN, yet still being a segment.

Table 3.9. Vowel alternations between different syllable structures in Wuning Gan (Sagart 1990, p. 151)

	V	Vʔ	VV	VN
/æ/	[æ]	[æʔ]	[ai] [jau]	[an] [jaŋ]
/œ/	[œ]	[œʔ]	[ɵi]	[ɵn]

Based on the reasoning for determining the status of glottal stop used in the above two examples (Shaoxing Wu and Wuning Gan), it is possible to use vowel change to argue for ʔ being a segment. If glottal stop patterns with other codas in terms of vowel distribution, it is more like to be a segment. However, vowel changes *cannot* be used as to argue for ʔ being a suprasegment definitively. If glottal stop does not pattern with other codas in terms of vowel distribution, it could either be due to ʔ being a suprasegment, or be due to ʔ, as a segment, lacking place feature (or lacking sonorancy in the case of Wuning Gan).

3.2.1.2 Evidence of glottal stop status from onset assimilation

The second perspective for discussing the status of glottal stop is from the patterning of onset assimilation. Onset assimilation that will be discussed in this section refers to the phenomenon that when two words are juxtaposed to each other, the onset of the second word is lenited in a way determined by the syllable structure of the preceding syllable. This phenomenon is frequently observed in the Min language family, and is referred to as *Shēngmǔ Lèihuà* 声母类化 ("onset assimilation") in Chinese literature. I have observed four types of onset assimilation pattern in Min languages: 1) Vʔ syllables prevent onset assimilations whereas open syllables license them; 2) Vʔ syllables license onset assimilation as open syllables; 3) some Vʔ syllables license onset assimilation; some Vʔ syllables prevent onset assimilation. I argue that among the three types of onset assimilation pattern, the first type demonstrate the glottal stop in Vʔ is a segment; the second type cannot determine the glottal stop status; the third type demonstrate that the glottal stops in Vʔ syllables preventing onset assimilation are segmental, and the glottal stops in Vʔ syllables licensing onset assimilation are suprasegmental.

3.2.1.2.1 Putian Jiangkou Min: Vʔ syllables prevent onset assimilations

Putian Jiangkou 莆田江口 Min has /-ʔ/ as the only possible obstruent coda. Onset assimilation has been observed in Putian Jiangkou 莆田江口 Min (Huang 2013). In the language, the obstruent onset of the second word is lenited into a sonorant or deleted when the first word is an open syllable (13a, b, c), but remains unchanged when the first word ends in glottal stop (13d, e, f). At the same time, the glottal stop of the first word is changed into /p/, /t/, or /k/ based on the place of articulation of second word's onset (13d, e, f).

(13)

	Underlying	Surface	Example
a.	∅ + /p-/	∅ + ∅	/tɒ 35 puai 533/ → [tɒ 11 uai 533] 茶杯 "tea cup"
b.	∅ + /t-/	∅ + [l-]	/kɔu 42 tai 11/ → [kɔu 35 lai 11] 古代 "ancient"
c.	∅ + /k-/	∅ + ∅	/ɬei 51 kai 51/ → [ɬei 55 ai 51] 世界 "world"
d.	/-ʔ/ + /p-/	[-p] + [p-]	/hɔʔ 24 pu 11/ → [hɔp 21 pu 11] 服务 "serve"
e.	/-ʔ/ + /t^h-/	[-t] + [t^h-]	/koʔ21 t^hau 35/ → [kot 24 t^hau 35] 骨头 "bone"
f.	/-ʔ/ + /k-/	[-k] + [k-]	/leʔ 24 k^hi 51/ → [lek 21 k^hi 51] 力气 "strength"

(Huang 2013)

As shown in Example (13), Vʔ behaves differently from open syllables in terms of the onset alternation pattern of the word following them. Glottal stop functions as a sufficient barrier to prevent the following onset from undergoing lenition, which can be evidence for the glottal stop in Vʔ being a segment. In addition, the glottal stop coda in the first word changes its place of articulation to that of the second syllable onset. Such alternations of glottal stop provides further support for the glottal stop being a segment.

3.2.1.2.2 Nantong Jianghuai: the [-ʔ] in Vʔ syllables is /-k/ underlyingly

Similar to Putian Jiangkou Min, Nantong Jianghuai also has the glottal stop as the only possible coda when the words are produced in isolation. The onset assimilation pattern in Nantong Jianghuai is also similar to Putian Jiangkou Min (Ao 1993). In the language, the stops and affricates in the onset of the second word become fricatives or tap when following an open syllable, but remains unchanged when following a Vʔ syllable. The onset assimilation pattern in V- and Vʔ-initial words in Nantong Jianghuai is shown in Table 3.10.

Table 3.10. Onset assimilation pattern V- and Vʔ-initial words in Nantong Jianghuai (Ao 1993).

Underlying	After [V]	After [Vʔ]
p	β	p
p^j	$β^j$	p^j
t	ɾ	t
t^j	$ɾ^j$	t^j
ts	z	ts
tʃ	ʒ	tʃ
tɕ	ʑ	tɕ
$tɕ^w$	$ʑ^w$	$tɕ^w$

Table 3.10 shows that the glottal stop in Vʔ syllables in Nantong Jianghuai functions as a barrier to the onset lenition. This can be evidence for the glottal stop as a segment in the language. In addition, Ao (1993) maintains that the obstruent coda in the surface form is /k/ underlyingly. The evidence is from the assimilation pattern of the glottal coda in disyllabic

compound words. As Example (14) shows, in a disyllabic compound, when the Vʔ is in the first position, the ʔ coda undergoes assimilation conditioned by the onset of the following syllable (14 a). When the onset of the second syllable is an obstruent or a sonorant (except for glide), the obstruent coda of the first syllable is assimilated into the same segment as second syllable's onset (14b, c). When the onset of the second syllable is a glide or a vowel, the obstruent coda of the first syllable is deleted (14d, e). However, as shown in Example (15), when the second syllable is a toneless word, the obstruent coda of the first syllable is not assimilated. When the coda assimilation is prohibited, the obstruent coda surfaces as [k] (15a, b, c). When the toneless word has a sonorant onset, the sonorant onset is also assimilated into [k] (15 b, c). Ao (1993) thus claims that /k/ is the underlying form of [ʔ], and proposes an additional rule that the [+dorsal] place feature surfaces only when the syllable is in non-final positions, which explains why monosyllabic words only have glottal stop obstruent in the surface.

(14)

	Manner	Underlying	Surface	Example
a.	Isolation	/-k/	[-ʔ]	/t^{h}**ik** HM/ → [t^{h}i**ʔ** HM] "iron"
b.	Obstruent	/-k/ + /p-/	[-p] + [p-]	/t^{h}**ik** HM + **p**ɔŋ H/ → [t^{h}**ip** HM + **p**ɔ̃ H] "iron plate"
		/-k/ + /t^{h}-/	[-t] + [t^{h}-]	/t^{h}**ik** HM + **t^{h}**e MH/ → [t^{h}**it** HM + **t^{h}**e MH] "iron head"
		/-k/ + /k^{hw}-/	[-k^{hw}] + [k^{hw}-]	/t^{h}**ik** HM + **k^{hw}**ɔŋ MH/ → [t^{h}**ik** HM + **k^{hw}**ɔ̃ MH] "iron ring"
c.	Liquid	/-k/ + /l-/	[-l] + [l-]	/t^{h}**ik** HM + **l**β̩ LM/ → [t^{h}**il** HM + **l**β̩ MLM] "railroad"
d.	Glide	/-k/ + /j-/	∅ + [j-]	/tsɛ**k** HM + **j**ə HM/ → [tsɛ HM + **j**ə HM] "excerpt"
e.	∅	/-k/ + ∅	∅ + ∅	/tsɛ**k** HM + ʒ̩ MH/ → [tsɛ HM + ʒ̩ MH] "inquire"

(Ao 1993)

(15)

	Manner	Underlying	Surface	Example				
a.	Obstruent	/-k/ + /s-/	[-k] + [s-]	/ʧɨ**k**	**s**e (toneless)/			
				[ʧɨ**k** HM	**s**e L]			
				eat	REQUESTIVE			
				Please eat.				
b.	Liquid	/-k/ + /l-/	[-k] + [k-]	/ŋu	y	pɛ	ʧʰɨ**k**	**l**o (toneless)/
				[ŋu H	y M	pɛ M	ʧʰɨ**k** HM	**k**o L]
				I	and	not	eat	RETORTIVE
				But I won't eat!				
c.	Glide	/-k/ + /j-/	[-k] + [k-]	/xə	ji**k**	**j**ɔ (toneless)/		
				[xə H	ji**k** H	**k**ɔ H]		
				well	hot	EXCLAMATIVE		
				How hot!				

(Ao 1993)

3.2.1.2.3 Gutian Pinghu Min: Vʔ syllables licences onset assimilations

There are also languages in which Vʔ syllables trigger onset lenition just as open syllables do. Gutian Pinghu 古田平湖 Min is an example (Chen 2006). The onset assimilation pattern of the second word in bimorphemic words is listed in Example (16). As shown in Example (16), when the syllable is open (16a, b, c) or closed by glottal stop (16d, e, f), the obstruent onset of the second word is either lenited to a fricative or a sonorant or deleted completely. When the syllable is closed by /k/ (16g, h, i), the onset of the second word does not change. Comparing open syllables, Vʔ, and Vk, we see that /k/ prevents the onset lenition in the second syllable, whereas ʔ licenses the lenition. Although Vʔ does not behave the same as Vk, it does not necessarily disqualify ʔ from being analyzed as a segment. We can appeal to the same reasoning – the placeless nature of glottal stop – as the discussion on the vowel quality alternation in Wuning Gan (Sagart 1990) and translaryngeal harmony in Zapotec (Section 2.1). Because glottal stop is placeless, it may not be a "strong" enough segment as other consonants (i.e. /k/) to prevent the lenition of the following onset.

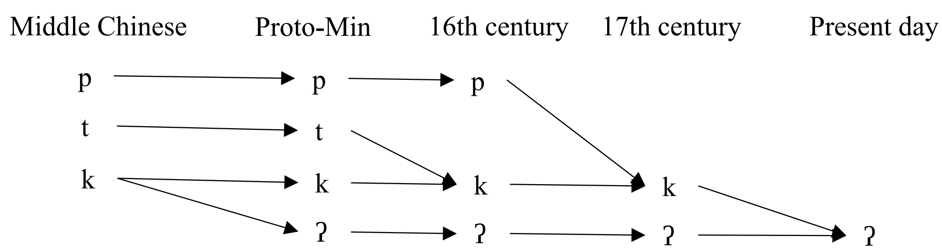
(16)	Underlying	Surface	Example
a.	∅ + /p-/	∅ + [β-]	/ta 33 puoi 55/ → [ta 33 βuoi 55] 茶杯 “teacup”
b.	∅ + /t-/	∅ + [l-]	/k^huo 55 tyoŋ 42/ → [k^huo 21 lyoŋ 42] 科长 “head of office”
c.	∅ + /k-/	∅ + ∅	/mi 42 kouŋ 55/ → [mi 21 ouŋ 42] 米缸 “rice container”
d.	/-ʔ/ + /p-/	∅ + [β-]	/tɔʔ 2 puo 21/ → [tɔ 55 βuo 51] 桌布 “table cloth”
e.	/-ʔ/ + /t-/	∅ + [l-]	/kuoʔ 2 tyoŋ 42/ → [kuo 33 lyoŋ 42] 局长 “head of bureau”
f.	/-ʔ/ + /k-/	∅ + ∅	/paʔ 5 kaʔ 2/ → [pa 21 aʔ 2] 白鸽 “white pigeon”
g.	/-k/ + /p-/	[-k] + [p-]	/ŋuok 5 piaŋ 42/ → [ŋuok 21 piaŋ 42] 月饼 “moon cake”
h.	/-k/ + /t-/	[-k] + [t-]	/nik 5 tau 21/ → [nik21 tau 21] 日昼 “noon”
i.	/-k/ + /k-/	[-k] + [k-]	/niek 2 kiaŋ 42/ → [niek 33 kiaŋ 42] 镊子 “tweezers”

(Chen 2006, p. 96)

3.2.1.2.4 Fuzhou Min: Some Vʔ syllables prevent onset assimilation; some license it.

There is a third type of onset assimilation pattern, in which some Vʔ words trigger onset lenition in the second syllable whereas the other Vʔ words do not. Fuzhou Min exhibits such onset assimilation pattern (Chan 1993, 1997). The only possible obstruent coda in Fuzhou Min is /ʔ/. However, the /ʔ/ historically comes from two sources. Proto-Min has four obstruent codas – /*-p, *-t, *-k, *-ʔ/. The glottal stop is derived from /*-k/ when preceeded by the low vowel /*a/. In 16th century Fuzhou Min, /*-t/ merged to /*-k/, and in 17th century, /*-p/ merged to /*-k/. In modern Fuzhou Min, /ʔ/ is the reflex of Proto-Min /*-k/ and /*-ʔ/. The diachronic change of Fuzhou Min from Middle Chinese to the current time is schematized in Example (17).

(17)



(Chan 1997, p. 152)

Despite the merger of /*-p, *-t, *-k, *-ʔ/ > /-ʔ/, we can still see still traces of the distinction between Proto-Min /*-k/ and /*-ʔ/ in the onset alternation and sandhi patterns (see Example (18)). In sequences of CVʔ+CV, for CVʔ syllables with ʔ< /*-k/, the onset of the second word does not change (18a, b); further, the glottal stop in the first word is preserved. For CVʔ syllables with ʔ< /*-ʔ/, the onset of the second word is deleted (18c, d). Further, the glottal stop in the first word is deleted. For open syllables, the onset of the following word is deleted, same as syllables closed by ʔ< /*-ʔ/ (18 e, f).

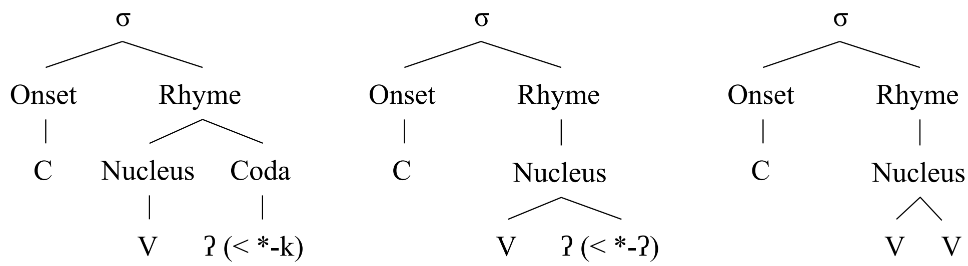
(18)

	Underlying	Surface	Examples
a.	/-ʔ (*-k)/ + /k-/	[-ʔ] + [k-]	/pøy**ʔ** **(<*-k)** 13 **k**iŋ 44/ → [pøy**ʔ** 2 **k**iŋ 44] “Beijing”
b.	/-ʔ (*-k)/ + /x-/	[-ʔ] + [x-]	/suo**ʔ** **(<*-k)** 13 **x**ua 44/ → [suo**ʔ** 2 **x**ua 44] “snowflake”
c.	/-ʔ (*-ʔ)/ + /k-/	∅ + ∅	/pa**ʔ** **(<*-ʔ)** 13 **k**uŋ 44/ → [pa 44 uŋ 44] “grandfather’s elder brother”
d.	/-ʔ (*-ʔ)/ + /x-/	∅ + ∅	/suo**ʔ** **(<*-ʔ)** 5 **x**ui 44/ → [suo 44 ui 44] “lime”
e.	∅ + /k-/	∅ + ∅	/tieŋ 131 po 213 **k**uoʔ 5/ → [tieŋ 22 mo 44 uoʔ 5] “telegraph office”
f.	∅ + /x-/	∅ + ∅	/tuai 131 **x**ouʔ 5/ → [tuai 22 ouʔ 5] “university”

(Chan 1997)

Chan (1997) provides a metrical analysis to this glottal stop-split phenomenon, as schematized in Example (19). Chan (1997) outlined that every syllable is consists of an onset and a rhyme. Within the rhyme, there is a nucleus and an optional coda. The glottal stop that is derived from /k/ occupies the coda position of the syllable (19a), whereas the glottal stop that is derived from Proto-Min /ʔ/ is part of the nucleus (19b). It occupies the same position as the second vowel in a diphthong (19c). This analysis also explains why diphthong does not occur before /ʔ/ (*-ʔ). Since the glottal stop occupies the second position in the nucleus, there is no slot for an additional vowel to form a diphthong. Based on this analysis, /-ʔ/ < /*-k/ is a segment because it is in the coda position; /-ʔ/ < /*-ʔ/ is a suprasegment because it is part of the nucleus. I adopt this analysis because in Fuzhou Min, in the phenomenon of licensing the onset lenition of the following syllable, syllables closed with /-ʔ/ < /*-ʔ/ behaves the same as open syllables, but differently from syllables closed with /-ʔ/ < /*-k/. We can no longer appeal to the “placeless” nature of glottal stop to argue for a segmental status for /-ʔ/ < /*-ʔ/, because /-ʔ/ < /*-k/ can function as a consonantal barrier against onset lenition, which indicates that being placeless does not prevent a segment from being as a lenition barrier. Thus, I suggest revising the transcription as /Vʔ/ (< /*-k/) and /Vˀ/ (< /*-ʔ/) for Fuzhou Min.

(19) a. ʔ (< *-k) b. ʔ (< *-ʔ) c. Diphthong



(C: consonant; V: vowel. Chan 1997, p. 279)

To summarize, the glottal stops in Vʔ in Chinese languages behave in two ways in the onset assimilation. On one hand, there are glottal stops that behave the same as other segment. For example, glottal stop in Shaoxing Wu behaves similarly as nasals in that it can be moraic and close syllables with lax vowels. And in Putian Jiangkou Min and Fuzhou Min, words following Vʔ rhymes and following Vʔ (< /*-k/) in Fuzhou Min do not trigger onset lenition, while open syllables do. On the other hand, there are examples of Vʔ that behaves similarly to open vowels. In Wuning Gan, vowels are centralized in nasal-closed syllables, but are not centralized in open syllables or vʔ syllables. Vʔ in Gutian Pinghu Min and Vʔ (< /*-ʔ/) in Fuzhou Min trigger the same type of onset lenition in the following syllable as open syllables in those languages do. Glottal stops in Vʔ syllables that behave similarly to open syllables can be analyzed as suprasegment (i.e. a glottalized phonation realized on vowel). However, we can also attribute the comparable behavior between Vʔ and open syllables to the placeness nature of the glottal stop segment (not for Fuzhou Min because /-ʔ/ < /*-k/ does prevent onset lenition).

3.2.1.3 Summary

In Sections 3.2.1.1 and 3.2.1.2, I review the evidence of /-ʔ/ as a segment and as a suprasegment from the perspectives of vowel quality and onset assimilation pattern. I summarize the evidence that can determine the status of /-ʔ/ in Table 3.11.

Table 3.11. Criteria for glottal stop being segmental and suprasegmental in Vʔ with examples from Chinese languages

ʔ status	Language	Reason	Source
segmental /Vʔ/	Shaoxing Wu	[Vʔ] shares the same vowel quality as [VN] and differs from [V].	Zhang 2006
	Putian Jiangkou Min	[Vʔ] prevents onset assimilation in [Vʔ +CV] whereas [V] licenses onset assimilation in [V + CV].	Huang 2013
	Nantong Jianghuai Mandarin	[ʔ] in [Vʔ] is /k/ underlyingly.	Ao 1993
	Fuzhou Min	[Vʔ (<*-k)] prevents onset assimilation in [Vʔ (<*-k) + CV], whereas [V] licenses onset assimilation in [V + CV]. The glottal stop in [Vʔ (<*-k)] is segmental.	Chan 1997
suprasegmental /V^{ʔ}/	Fuzhou Min	[Vʔ (<* -ʔ)] licenses onset assimilation in [Vʔ + CV], whereas [Vʔ (<*-k)] prevents onset assimilation in [Vʔ (<*-k) + CV]. The glottal stop in [Vʔ (<*-ʔ)] is suprasegmental.	Chan 1997

3.3 Reflexes of **Rù* tone in Modern Chinese

In Section 3.1, I mentioned that **Rù* syllables are associated with two **Rù* tones – **Yīnrù* and **Yángrù* in the later stage of Middle Chinese. In this section, I discuss the reflexes of these two **Rù* tones in Middle Chinese.

I categorize the 98 language varieties in the survey into four categories by their reflex pattern of the Middle Chinese **Rù* tone. Languages of Type I have more than two **Rù* tone reflexes. The increase in the number of **Rù* tone in the reflex is due to a later binary register split of Yīnrù and/or Yángrù. The split categories within Yin and Yang registers are called "Upper *Yīnrù*", "Lower *Yīnrù*", "Upper *Yángrù*", and "Lower *Yángrù*" (i.e., 上阴入; 下阴入; 上阳入; 下阳入) tones in the Chinese linguistic tradition . The split of **Rù* into upper and lower registers is non-arbitrary in some languages. The split could be based on the onset, the coda, or the vowel quality. Examples of the **Rù* tone split are in Table 3.12.

Table 3.12. **Rù* tone splits in Modern Chinese

Language	Tone	Environment
Duchang Gan	*Upper *Yīnrù* & *Yángrù*	Voiceless, non-/h/ onsets
	*Lower *Yīnrù* & *Yángrù*	Voiced or /h/ onsets
Anyi Gan	*Upper *Yīnrù*	/-p/ or /-t/ codas
	*Lower *Yīnrù*	/-ʔ/ coda
Cantonese	*Upper *Yīnrù*	short vowels
	*Lower *Yīnrù*	long vowels
Dongkeng Yue	*Upper *Yīnrù*	/ak, uak, iœk, iɛk, ek, ɔk, ɤk, iœk, uɤk/
	*Lower *Yīnrù*	/aʔ, uaʔ, iœʔ, iɛʔ, eʔ, ɔʔ, ɤʔ, iœʔ, uɤʔ/
Tangxia Yue	*Lower *Yīnrù*	/aʔ, iæ^{a}ʔ, uɔʔ/
	**Yángrù*	/ak, iæ^{a}k, uɔk/

Languages of Type II preserve the two **Rù* tones categories in Middle Chinese: **Yīnrù* and **Yángrù*. Languages of Type III have only one **Rù* tone in the reflex, because the *Yīn* and *Yáng* distinction in **Rù* tone disappears. Languages of Type IV do not have any reflexes of the **Rù* tones. Middle Chinese **Rù* tones are merged with **Píng*, **Shǎng*, and/or **Qù* tones. Within each type, I further categorize the language varieties into two sub-types: Sub-type A: syllables bearing **Rù* tone have obstruent codas (/p, t, k, ʔ/ or a subset of them); Sub-type B: syllables bearing **Rù* tone do not have obstruent codas. I list examples of each type of **Rù* tone and **Rù* syllable reflex in Table 3.13.

Table 3.13. Reflexes of Middle Chinese **Rù* tone in Modern Chinese.

Tone	Rhyme	Language	Reflexes of **Rù* tone	Examples	Checked syllable?	Checked tone?
I. More than two **Rù* tones	A. Have coda	Duchang 都昌 Gan	Upper **Yīnrù* 45 Lower **Yīnrù* 24 Upper **Yángrù* 3 Lower **Yángrù* 21	/tal 45/<*/tʌp/ 答 "answer" /ləl 24/<*/ɳiɪp/ 入 "enter" /ʂ̥ɛl 3/<*/ɦɣuat/ 舌 "tongue" /lal 21/<*/tʰʌp/ 踏 "step"	Yes	Yes

continued …

Table 3.13. Reflexes of Middle Chinese **Rù* tone in Modern Chinese (continued.)

Tone	Rhyme	Language	Reflexes of **Rù* tone	Examples	Checked syllable?	Checked tone?
		Anyi 安义 Gan	**Yīnrù*A 5 **Yīnrù*B 53 **Yángrù* 2	/tɔp 5/<*/tʌp/ 答 /lɤp 5/<*/ȵiɪp/ 入 /pɔʔ 53/<*/pɑk/ 博 "broad" /sɛt 2/<*/ɦɣuat/ 舌 /tʰap 2/<*/tʰʌp/ 踏	Yes	Yes
		Hong Kong Cantonese	Upper **Yīnrù* 5 Lower **Yīnrù* 33 **Yángrù* 2	/jek 5/<*/ʔiᴇk/ 益 "benefit" /jak 33/<*/kʰek/ 喫 "eat" /jek 2/<*/jiᴇk/ 亦 "also"	Yes	Yes
		Dongkeng 东坑 Yue	Upper **Yīnrù* 5 Lower **Yīnrù* 23 **Yángrù* 3	/kʌk 5/<*/kɣiɪp/ 急 "hurry" /kaʔ 23/<*/kʌp/ 鸽 "pigeon" /ʃʌk 3/<*/dʑiɪp/ 十 "ten"	Yes	Yes
		Tangxia 塘厦 Yue	Upper **Yīnrù* 5 Lower **Yīnrù* 13 **Yángrù* 3	/kɐp 5/<*/kɣiɪp/ 急 /ʃik 13/<*/siuᴇt/ 雪 "snow" /tik 3/<*/dep/ 叠 "fold"	Yes	Yes
II. Two **Rù* tones	A. Have coda	Wengyuan 翁源 Hakka	**Yīnrù* 2 **Yángrù* 5	/tak 2/<*/tʌp/ 答 /ŋit 5/<*/ȵiɪp/ 入 /sat 5/<*/ɦɣuat/ 舌 /tʰak 5/<*/tʰʌp/ 踏	Yes	Yes
	B. No coda	Yueyang (Xinqiang) 岳阳 (新墙) Gan	**Yīnrù* 5 **Yángrù* 3	/xe 5/<*/həkl/ 黑 "black" /xø 3/<*/ɦʌp/ 盒 "box"	No	No
III. One **Rù* tone	A. Have coda	Ganxian (Panlong) 赣县 (蟠龙) Gan	**Rù* 5	/taʔ 5/<*/tʌp/ 答 /iɛʔ 5/<*/ȵiɪp/ 入 /sɛʔ 5/<*/ɦɣuat/ 舌 /tʰaʔ 5/<*/tʰʌp/ 踏	Yes	Yes
	B. No coda	Liling (Baitutan) 醴陵 (白兔潭) Gan	**Rù* 45	/ta 45/<*/tʌp/ 答 /i 45/<*/ȵiɪp/ 入 /se 45/<*/ɦɣuat/ 舌 /tʰa 45/<*/tʰʌp/ 踏	No	No
IV. No **Rù* tone	B. No coda	Changting 长汀 Hakka	∅	/ta 24 Yángpíng/<*/tʌp/ 答 /ne 21 Yánqù/<*/ȵiɪp/ 入 /ʃe 21 Yánqù/<*/ɦɣuat/ 舌 /tʰa 21 Yánqù/<*/tʰʌp/ 踏	No	No

As shown in Table 3.13, for Type I languages, which have more than two **Rù* tones in the modern forms, the reflexes of the **Rù* tones are all borne by obstruent-closed syllables. The survey did not find a variety that has more than one **Rù* tones but where the obstruent coda in **Rù* syllable is lost. For languages of Type II (have **Yīnr'u* and **Yángrù* tones) and Type

III (have one *Rù tone), the reflexes of *Rù* tone(s) can be carried by syllables with or without obstruent codas. For varieties whose **Rù* tones are borne by open or nasal-closed syllables, the obstruent codas in Middle Chinese **Rù* syllables are lost, but there are distinct tonal contours that are historically associated with the Middle Chinese **Rù* syllables. If we look at the tone system of those varieties only from a synchronic perspective, syllables derived from **Rù* tones are not distinct from those that are not (except for the difference in tonal values). Languages of Type IV do not have any distinct category that is a reflex of the **Rù* tones, nor do they have syllables that are closed by obstruent codas. In summary, based on the survey results, there are languages that have distinct categories as reflexes of the **Rù* tone but do not have obstruent-closed syllables. However, there are no languages that have obstruent-closed syllables and where those obstruent-closed syllables are *not* marked by distinctive tones from other syllables.

3.3.1 Are tonal reflexes of **Rù* tones allophonic?

In some languages, the reflexes of **Rù* tones have the same tonal value as the reflexes of non-**Rù* tones. Several studies find that the reflexes of **Rù* tones overlap with the reflexes of non-**Rù* tones in F0 (e.g., Cantonese: Chan 1987; Fuzhou Min: Shao 2012; Taiwanese Min: Pan 2017; Kuo 2013). In such cases, are those reflexes of **Rù* tones allotones of non-**Rù* tones? Some researchers analyze the reflexes of **Rù* tones as the allotones of the reflexes of non-**Rù* tones based on their similar F0 values. For example, Tan (2016) proposes there are eight tones in Taishan Yue, the reflexes of **Rù* tones (45, 33, 42) are the shorter versions of **Yīnshǎng* (45), **Yīnpíng* (33), and **Qù* (42) tones[3]. Here, I consider that phonetic evidence alone does not provide a definitive argument for reflexes of **Rù* tones being treated as allotones of non-**Rù* tones. Phonological evidence showing that the reflexes of **Rù* tones are derived from non-**Rù* tones (e.g., through alternations) is stronger support in favor of treating them as allophonic.

3. Cheng (1973) maintains that there are five tones in Taishanese Yue: 66, 44, 22, 52, and 31; all but 22 can be borne by stop-closed syllables. The discrepancy between Cheng (1973) and Tan (2016) might be due to the diachronic sound change of the language.

3.3.1.1 The High **Rù* tone in Nantong Jianghuai as allotone of Low-Mid tone

Some studies have provided phonological evidence that the reflexes of **Rù* tone is derived synchronically from the reflexes of non-**Rù* tone in the language. Nantong Jianghuai Mandarin has five tones that are the reflexes of non-**Rù* tones: Low (L), High (H), Low-Mid (LM), Mid-High (MH), High-Mid (HM), and two tones that are the reflexes of **Rù* tones: H and HM (referred to as H (**Rù*) and HM (**Rù*) hereafter). Ao (1993) proposes that the H (**Rù*) tone is a LM non-**Rù* tone underlyingly, because the H (**Rù*) tone shares the same sandhi pattern as the LM non-**Rù* tone. Table 3.14 shows the sandhi pattern when the first syllables are of Tones LM, H, H (**Ru*) followed by sonorant onset, and H (**Ru*) followed by obstruent onset. Table 3.14 shows that H (**Ru*) shares the same sandhi pattern as LM when the onset of the following syllable is a *sonorant*, but the same tone shares the same sandhi pattern as H when the onset of its following syllable is an *obstruent*. Example (20) shows examples of the sandhi pattern when the first syllable has the tones of LM, H (**Ru*) followed by sonorant onset, and H (**Ru*) followed by obstruent onset. Ao (1993) did not provide examples of the sandhi pattern when the first syllable has the tone of H.

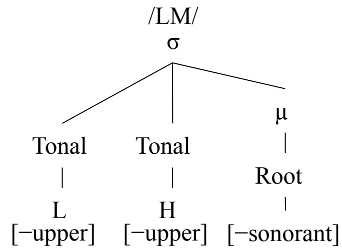
Table 3.14. Disyllabic sandhi rules when the first syllables are of Tones LM, H, H (**Ru*) followed by sonorant onset, and H (**Ru*) followed by obstruent onset.

2nd σ / 1st σ	LM	L	HM	MH	H	H (**Ru*)
LM	MH + MLM	MH + ML	MH + HM	M + MH	M + H	M + H
H	H + MLM	H + ML	H + HM	H + MH	H + H	H + H
H (**Ru*) (+sonorant onset)	MH + MLM	MH + ML	MH + HM	M + MH	M + H	M + H
H (**Ru*) (+obstruent onset)	H + MLM	H + ML	H + HM	H + MH	H + H	H + H

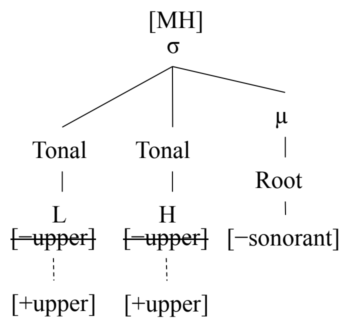
(20)	Sandhi pattern	Examples
a.	LM + H → M + H	/t^{h}iŋ LM + jɨŋ H/ → [t^{h}ĩ M + jiŋ H] “electric image (movie)”
b.	LM + H (**Rù*) → M + H	/t^{h}iŋ LM/ + lik H → [t^{h}ĩ M + liʔ H] “electric power”
c.	LM + HM → MH + HM	/t^{h}iŋ LM + ɕiŋ HM/ → [t^{h}ĩ MH + ɕĩ HM] “electric wire”
d.	H (**Rù*) + H (sonorant onset) → M + H	lok H + /ʋβ̩ H/ → [lo M + ʋβ̩ H] “six-five”
e.	H (**Rù*) + H (**Rù*, sonorant onset) → M + H	lok H + lok H → [lol M + loʔ H] “six-six”
f.	H (**Rù*) + HM (**Rù*, sonorant onset) → MH + HM	lok H + /jik HM/ → [lo MH + jiʔ HM] “six-one”
g.	H (**Rù*) + H (obstruent onset) → H + H	lok H + /ʨy H/ → [lot H + ʨy H] “six-nine”
h.	H (**Rù*) + H (**Rù*, obstruent onset) → H + H	lok H + sɛk H → [los H + sɛʔ H] “six-ten”
i.	H (**Rù*) + HM (**Rù*, obstruent onset) → H + HM	lok H + /pɔk HM/ → [lop H + pɔʔ HM] “six-eight”

Ao (1993) argues that the underlying tone of H (**Ru*) is /LM/. The tone surfaces as [H] in isolation forms and when the following syllable has an obstruent coda, because in those two circumstances, the tone is associated with syllables closed by obstruent codas. The presence of obstruent coda causes a tonal change of /LM/. As discussed in Section 3.2.1.2.2, in Nantong Jianghaui Mandarin, the reflexes of **Ru* syllables are closed by /k/ underlyingly. As shown in Example (14), when the reflexes of **Ru* syllables are in isolation, their obstruent codas surface as [ʔ]. When /Vk/ is followed by syllables with obstruent codas, the coda /k/ remains to be an obstruent. When /Vk/ is followed by syllables with sonorant codas, the coda /k/ is assimilated to a sonorant. Example (21) illustrates how /LM/ tone borne by obstruent-closed syllables becomes a [H] tone in the surface form. Ao (1993) proposed that the /LM/ tone is /LH/ tone with a [-upper] register (21a). When the syllable bearing /LM/ tone is closed by an obstruent, /LM/ goes through two steps of changes. First, the obstruent coda causes an insertion of [+upper] register. The tone becomes [MH] after the [+upper] register insertion (21b). Second, the first tonal node is deleted (21c). The surface form of the tone becomes [H] (21d). In contrast, when the codas of /Vk/ syllables become sonorants, the tonal contour of the /LM/ tone is not affected by the coda. Thus, their tonal pattern is the same as the tone /LM/ associated with open syllables and sonorant-closed syllables.

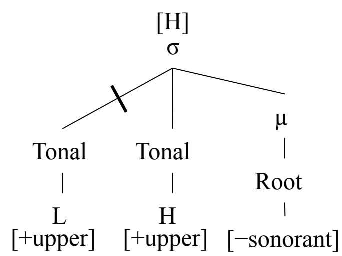
(21) a. Underlying form



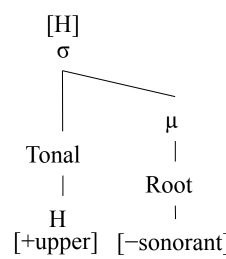
b. [+upper] register insertion



c. First tonal node deletion



c. Surface form



3.3.1.2 Zhangzhou Min: 41 <**Yīnrù* and 2 <**Yángrù* as different tonemes from 41 <**Yīnqù* and 22 <**Yángpíng*

Huang (2018) proposes that Zhangzhou Min has eight tones: 35 < **Yīnpíng*, 22 <**Yáng-píng*, 51 <**Shǎng*, 41 <**Yīnqù*, 33 <**Yángqù*, 41 <**Yīnrù*, 221 <**Yángrù*, and 22 <**Yángrù*. Tones 41 <**Yīnrù* and 221 <**Yángrù* only occurs with syllables ending with /-p, -t, -k/ codas. Tone 22 <**Yángrù* only occurs with open syllables. Huang (2018) maintains that 41 <**Yīnrù* is a different toneme than 41 <*_Yīnqù_, and 22 <*_Yángrù_ is a different phonemic tone from 22 <*_Yángpíng_.

Huang (2018) maintained that 41 <*_Yīnrù_ is a different tone from 41 <*_Yīnqù_ because they differ in duration, voice quality, vowel quality, coda type, and tonal value in sandhi forms. In terms of duration, 41 <*_Yīnqù_ is longer than 41 <*_Yīnrù_. In terms of voice quality, high vowels with 41 <*_Yīnqù_ is breathy whereas high vowels with 41 <*_Yīnrù_ is creaky. In terms of vowel quality, high vowel is diphthongized when it bears 41 <*_Yīnrù_, but does not go through diphthongization when it bears 41 <*_Yīnqù_. In sandhi forms, 41 <*_Yīnqù_ becomes a falling tone 63 whereas 41 <*_Yīnrù_ becomes an extra high-falling tone 65. I do not regard the above differences between 41 <*_Yīnqù_ and 41 <*_Yīnrù_ as definitive evidence for them being different tonemes. First, the difference in duration and voice quality could be a by-product of the syllable structure difference between the syllables that bear 41 <*_Yīnqù_ and 41 <*_Yīnrù_. Tone 41 <*_Yīnqù_ occurs in open or nasal-closed syllables, whereas 41 <*_Yīnrù_ occurs in obstruent-closed syllables. Across languages, we generally see that obstruent-closed syllables are shorter than open and nasal-closed syllables (Farnetani and Kori 1986; Rietveld and Frauenfelder 1987; Santen 1992; Maddieson and Ladefoged 1985). The creaky voice quality associated with 41 <*_Yīnrù_ could be induced by the occlusive glottal gesture associated with the obstruent coda of the syllable. The diphthongization observed in 41 <*_Yīnrù_ is also likely to be motivated by the syllable structure because high vowels are also diphthongized when bearing 221 <*_Yángrù_. It is more plausible that the diphthongization is specific to obstruent-closed syllables, rather than being tied to 41 <*_Yīnrù_. The only dimension that appeals to the property of tone itself is the difference between 41 <*_Yīnqù_ and 41 <*_Yīnrù_ in sandhi form. Tones 41 <*_Yīnqù_ and 41 <*_Yīnrù_ become 63 and 65 respectfully. However, it is worth considering that the sandhi form of 41 <*_Yīnrù_ being 65 rather than 63 is due to the extra short duration of the closed syllables. It is possible that *_Yīnrù_ does not have enough duration to reach the tonal target that is as low as *_Yīnqù_. Based on the above analysis, I argue that the evidence provided by Huang (2018) for treating 41 <*_Yīnqù_ and 41 <*_Yīnrù_ as distinct tonemes is not strong enough to be conclusive.

In contrast, their argument that 22 <_*Yángrù_ is distinct from 22 <_*Yángpíng_ is stronger. Tone 22 <_*Yángrù_ is borne by open syllables; it corresponds to the _*Yángrù_ tone borne by syllables reconstructed as closed by glottal stop, which has disappeared in modern-day Zhangzhou Min. Huang (2018) argued that 22 _*Yángrù_ and 22 _*Yángpíng_ are different tones because they have different tonal values in sandhi forms. Tone 22 <_*Yángrù_ becomes mid-falling 32 whereas 22 _*Yángpíng_ becomes mid-level 33 in sandhi forms. Their duration in sandhi forms is comparable. Thus, although 22 <_*Yangrù_ and 22 <_*Yangpíng_ appear to have been neutralized in their citation forms, their differences are preserved in sandhi forms.

With the examples of Nantong Jianghuai Mandarin and Zhangzhou Min, we see that tones derived from _*Rù_ tone can have the same F0 as tones derived from non-_*Rù_ tones, but still be a distinct toneme, whereas tones derived from _*Rù_ tone can also be an allotone of tones derived from non-_*Rù_ tone that have different F0 from them. Thus, it is not accurate to conclude a _*Rù_ tone as the allotone of a non-_*Rù_ tone just by solely looking at the F0. In order to justify whether a tone derived from _*Rù_ tone is an allotone of a tone derived from non-_*Rù_ tone, we need stronger support from phonological patterning showing that one tone can be derived from the other through a synchronic process. Before such phonological evidence is found, the relation between the reflexes of _*Rù_ tone and the reflexes of non-_*Rù_ tone remains to be ambiguous. Despite the ambiguous phonemic status of _*Rù_ tone, I confirm that, for all languages included in the survey, there are always fewer tones derived from _*Rù_ tone than tones derived from non-_*Rù_ tones. Even if we assume that all reflexes of _*Rù_ tones are allotones of the reflexes of non-_*Rù_ tones, we can still say that _*Rù_ allotones only occur for a subset of the non-_*Rù_ tones.

3.4 What does "checked" mean in Modern Chinese?

In Sections 3.2 and 3.3, I show that in many Chinese languages, there are still remnants of Middle Chinese _*Rù_ syllable and _*Rù_ tone, through reflexes that maintain a category distinction from other tones/syllables. The question that follows is: What is the relation between Middle

Chinese **Rù* syllable and **Rù* tone and Modern Chinese checked syllable and checked tone? Here I restate the definition of **Rù* syllable, **Rù* tone, checked syllable, and checked tone as following:

- **Rù* syllable: Rhymes that are reconstructed as closed by /*-p, *-t, *-k/ in Middle Chinese;
- **Rù* tone: Tones that are reconstructed as accompanying **Rù* syllables in Middle Chinese;
- Checked syllable: Syllables that are closed by obstruents and associated with distinctive tones from open syllables and sonorant-closed syllables, or a subset of tones borne by open syllables and sonorant-closed syllables;
- Checked tone: Tones that are borne by checked syllables.

The definitions of **Rù* syllables and **Rù* tones are very similar to those of checked syllables and checked tones. Their main difference is whether they are modeled reconstructions of Proto-Middle Chinese, or represent syllables in modern-day daughter languages. **Rù* syllable and **Rù* tone refer to a type of syllable and tone in Middle Chinese. The concept "checked" refers to a specific phonological category that combines syllable structure and tone together. "Checked" is a synchronic concept. According to the reconstructions of Middle Chinese, all **Rù* syllables and **Rù* tones were checked syllables and checked tones. The main difference is proto forms vs. synchronic ones. The reflexes of **Rù* syllable and **Rù* tone in Modern Chinese are no longer equivalent to the checked syllables and checked tones in the daughter languages. According to the survey, in Modern Chinese, the majority of checked syllables are derived from **Rù* syllables in Middle Chinese, and the majority of checked tones are derived from **Rù* tones in Middle Chinese. However, Modern Chinese words with checked syllables and checked tones in Modern Chinese are not in a one-on-one correspondence with **Rù* syllables and **Rù* tones in Middle Chinese. There exist reconstructed open syllables or nasal-closed syllables whose reflexes are obstruent-closed and bear tones that are designated to obstruent-closed syllables. Such words are not of **Rù* syllables or **Rù* tones in Middle Chinese, but synchronically count as checked syllable and checked tone in Modern Chinese. In Chinese linguistics, the phenomenon whereby non-**Rù* syllables acquire obstruent codas is called "*Shūshēng cùhuà* 舒声促化 (non-**Rù* syllable closing)" (Zhang 2006; Sun and Wang 2006; Du 2021). Conversely, there are reflexes of **Rù* syllables and tones in which the (reconstructed) obstruent coda was lost and have acquired tones that only occur in open or sonorant-closed syllables. In Chinese linguistics, the phenomenon whereby **Rù* syllables lost obstruent coda is called "*Rùshēng shūhuà* 入声舒化 (**Rù* syllable opening)" (Cao 2002; Feng 2011; Gu 2015; Shen 2007; Song 2009; Xing and Meng 2006; Yang 1982). Examples of non-**Rù* syllables in Middle Chinese acquiring obstruent coda in Modern Chinese is in (22). Examples of **Rù* syllables in Middle Chinese that have lost their obstruent coda in Modern Chinese are in (23).

(22)

Word	Modern Xinzhou Jin	Middle Chinese
去 "go"	kəʔ 2	*k^{h}ɨʌ *Shǎng*
祈 "pray"	ʨ^{h}iəʔ 2	*gɨi *Píng*

(Zhang 2006, p. 55)

(23)

Word	Modern Yiwu Wu	Middle Chinese
搭 "match"	ɗɔ 334	*tʌp *Rù*
月 "moon"	ȵye 213	*ŋʉɐt *Rù*
急 "hurry"	ʨiəʔ 4	*kɣiɪp *Rù*
夺 "take away"	dəʔ 1	*duat *Rù*

(Shi 2012, p. 85, 88)

To illustrate some general correspondences between checked syllable and tones, and the reflexes of Middle Chinese **Rù* syllable and **Rù* tone in Modern Chinese, I list the checked syllables, checked tones, **Rù* syllable reflex, and **Rù* tone reflex of 10 languages in Table 3.15, which include the same set of languages as Table 3.13. As Table 3.15 shows, I have yet to find a language that preserves the obstruent coda of **Rù* syllable in Middle Chinese but loses the **Rù* tone associated with the **Rù* syllable. There do, however, exist languages that have lost the obstruent coda of **Rù* syllable, but still preserves the **Rù* tone associated with the

Rù* syllable. I do not consider languages that belong to the latter category as having checked syllables or checked tones. Those languages do not have checked syllables because they do not have obstruent-closed syllables. Nor do I consider the tonal reflexes of **Rù* tone in those languages as checked tones, because synchronically it is impossible to determine whether the **Rù* tone reflexes belong to a different category from the non-Rù* tone reflexes. Tones that are derived from Middle Chinese **Rù* tone behaves the same as tones that are derived from Middle Chinese non-**Rù* tones, and all tones can be borne by the same syllable structures in the language.

To summarize, if we say a language has checked syllables and checked tones, it means the language has obstruent-closed syllables and the obstruent-closed syllables are associated with specific tone(s) synchronically. If we say a word in a language is of **Rù* syllable or **Rù* tone, it means that the word is closed by /p, t, k/ in Middle Chinese, and is associated with tones that are specific to /p, t, k/-closed syllables in Middle Chinese. Defining a word as “checked” makes reference to the word’s synchronic form, whereas identifying a word as “Rù” refers to the diachronic origin of the word in Middle Chinese phonology.

Table 3.15. Comparison of checked syllables, checked tones, reflex of **Rù* syllable, and reflex of **Rù* tone

Language	Checked syllable	Checked tone	Reflexes of **Rù* syllable	Categories of **Rù* tone
Duchang 都昌 Gan	/-k, -l/	45, 24, 3, 21	/k, l/	Upper **Yīnrù* 45 Lower **Yīnrù* 24 Upper **Yángrù* 3 Lower **Yángrù* 21
Anyi 安义 Gan	/-p, -t, -ʔ/	5, 53, 2	/-p, -t, -ʔ/	**Yīnrù*A 5 **Yīnrù*B 53 **Yángrù* 2
Hong Kong Cantonese	/-p, -t, -k/	5, 33, 2	/-p, -t, -k/	Upper **Yīnrù* 5 Lower **Yīnrù* 33 **Yángrù* 2
Dongkeng 东坑 Yue	/-k, -ʔ/	5, 23, 3	/-k, -ʔ/	Upper **Yīnrù* 5 Lower **Yīnrù* 23 **Yángrù* 3
Tangxia 塘厦 Yue	/-p, -t, -k, -ʔ/	5, 13, 3	/-p, -t, -k, -ʔ/	Upper **Yīnrù* 5 Lower **Yīnrù* 13 **Yángrù* 3
Wengyuan 翁源 Hakka	/-t, -k/	2, 5	/-t, -k/	**Yīnrù* 2 **Yángrù* 5
Yueyang (Xinqiang) 岳阳 (新墙) Gan	∅	∅	∅	**Yīnrù* 5 **Yángrù* 3
Ganxian (Panlong) 赣县 (蟠龙) Gan	/-ʔ/	5	/-ʔ/	**Rù* 5
Liling (Baitutan) 醴陵 (白兔潭) Gan	∅	∅	∅	**Rù* 45
Changting 长汀 Hakka	∅	∅	∅	∅

3.4.1 *Rùshēng shūhuà* (**Rù* syllable opening) in Modern Chinese

Reviewing the reflexes of **Rù* syllables and **Rù* tones in Modern Chinese, we can see that the *Rùshēng shūhuà* (**Rù* syllable opening) is a common tendency in Chinese languages, based on the reduction, debuccalization, and even loss of /*-p, *-t, *-k/ and the merger of **Rù* tone with non-**Rù* tones. As Sections 3.2 and 3.3 have shown, the reflex of **Rù* syllable and **Rù* tone has great variation. Such variation reflects the fact that **Rù* syllables and tones did not disappear completely all at once. Rather, the disappearance proceeded in multiple stages. Different varieties have gone through these stages in a different order and at a different rate. Zhu et al. (2008) listed three major paths leading to the loss of **Rù* syllables in Chinese languages. I schematize the different stages of each path proposed by Zhu et al. (2008) in Figure 3.1. The differences in the three paths depend on the sequencing of vowel lengthening, oral coda lenition and glottal coda loss. Path I involves syllable lengthening first, followed by oral obstruent codas lenition to glottal stop, and ending in the loss of the glottal coda. Path II first involves oral obstruent lenition to glottal stop first, followed by the loss of the glottal coda, and ending by syllable lengthening. Path III involves oral obstruent lenition to glottal stop first, followed by syllable lengthening, and ending in the loss of the glottal coda.

Looking at the different stages of **Rù* syllable "opening", at what point would the **Rù* syllable stop qualifying as a checked syllable using the criteria adopted here? Is there a stage before Stage 4 for which we could already say that "checkedness" has been lost in the **Rù* syllable? I propose that once the obstruent coda in **Rù* syllable is lost, the syllable is no longer checked. In Figure 3.1, the only stage before Stage 4 where checkedness has been lost is Stage 3 in Type II (Short V <**Rù*). In other cases, when the oral codas of **Rù* syllables are lenited to glottal stop, or when the vowels in **Rù* syllables are lengthened, the syllable is still an obstruent-closed syllable, which mean they still fit the definition of checked syllables. However, when the glottal stop coda of **Rù* syllable is lost, it is indistinguishable from unchecked open and nasal-closed syllables. Even if the syllable remains short, and is associated with a specific tone, we can posit that the language has a specific tone that happens to be short, and so it is unnecessary to still appeal to the concept of “checked”. An example of glottal coda being lost whereby the syllable’s short duration and tone are preserved is Nanjing Jianghuai Mandarin (Oakden 2017). Oakden (2017) reported that Nanjing Jianghuai Mandarin used to have /-ʔ/ as the reflex of *_Rù_ syllables but might have lost the /-ʔ/ coda based on the acoustic measurements of the vowels in /Vʔ/ syllables. Nanjing Jianghuai Mandarin has five tones. Four are derived from non-*_Rù_ tones: 31 <*_Yīnpíng_, 24 <*_Yángpíng_, 11 <*_Shǎng_, and 44 <*_Qù_. The other tone is 5 <*_Rù_, and occurs in syllables ending in /-ʔ/. Tone 5 <*_Rù_ has a significantly shorter duration than non-*_Rù_ tones. Additionally, vowels with Tone 5 <*_Rù_ have lower jitter values than those with the tones derived from non-*_Rù_ tones, suggesting that Tone 5 syllables are more modal than those derived from with non-*_Rù_ syllables. The modal quality in the syllables derived from *_Rù_ syllables may indicate a loss of glottal stop (which is generally expected to cause an _increase_ in jitter), despite the preservation of the short duration of the rhyme. Therefore, based on the definition of checked syllable and tone in this dissertation, Nanjing Jianghuai Mandarin does not have checked syllables or tones; we can reanalyze the tone 5 <*_Rù_ as a short high tone in the language, despite it being the reflex of a *_Rù_ tone. A similar case is Deqing Wu. Zhu and Jiao (2011) reported that Deqing Wu used to have syllables ending in /-ʔ/ as the reflex of the *_Rù_ syllables, but might have lost the /-ʔ/ in its modern variety. Deqing Wu has six tones derived from non-*_Rù_ tones (44 <*_Yīnpíng_, 213 <*_Yángpíng_, 53 <*_Yīnshǎng_, 31 <*_Yángshǎng_, 435 <*_Yīnqù_, 224 <*_Yánqù_) and two *_Rù_ tones (55 <*_Yīnrù_, 33 <*_Yángrù_). The two tones derived from *_Rù_ tone only occur in /-ʔ/-closed syllables. However, based on visual inspection of the spectrogram, Zhu and Jiao (2011) propose that the glottal coda in *_Rù_ tone is lost in Deqing Wu. The only difference between syllables derived from *_Rù_ syllables and those that are not is that the former are shorter. By the definition of checked syllable and tone, we will conclude that Deqing Wu does not have checked rhymes or checked tones.

There might also be an intermediate stage before the stage where the glottal stop derived from /*-p, *-t, *-k/ in *_Rù_ syllables (i.e. Vʔ) is lost completely. At this intermediate stage, the historical glottal stop becomes glottalization, which is a suprasegmental feature of the nucleus. At this stage, I determine that the language has lost checked syllables, but acquired a checked phonation. If the glottal stop in Vʔ is a suprasegment, it is similar to the checked phonation found in Zapotec languages. However, since rearticulated phonation is rarely reported in Chinese language, the odds that the checked phonation is contrastive with another glottalized phonation in the language are low. An example of this intermediate stage is syllable closed by /ʔ < *ʔ/ in Fuzhou Min. As I discussed in Section 3.2.1.2.4, the /ʔ< *ʔ/ in Fuzhou Min is a suprasegment, and syllables ended in /ʔ < *ʔ/ should be analyzed as /$V^ʔ$/. The /$V^ʔ$/ in Fuzhou Min is in fact at the intermediate stage before the complete loss of glottal stop in the language. The /$V^ʔ$/ in Fuzhou Min can accordingly be treated as having a checked phonation.

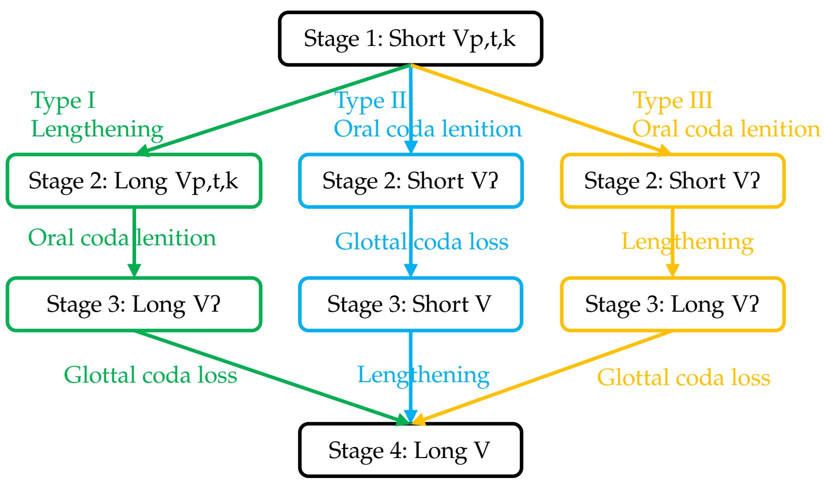


Figure 3.1. Three paths involving the loss of checkedness (schematized based on Zhu et al. 2008)

The distinction between **Rù* syllables and tones, and checked syllables and tones also speaks to the necessity of having a category of "checked" in the synchronic phonology of Chinese. "Checked" in synchronic Chinese is a phonotatic restriction between tone and syllable. This phonotactic constraint is rather homogenous among numerous varieties across all subfamilies of Chinese language. Such a homogenity is due to its historical origin in **Rù* syllables and **Rù* tones in Middle Chinese. At the same time, Middle Chinese **Rù* syllables and **Rù* tone are not strictly preserved in Modern Chinese. Thus, assigning a new name "checked" to describe this phonotactic constraint can be effective for identifying this restriction in the synchronic form of languages. The reason why the name "checked" is chosen is that the phonetic characteristics of checked syllables and tones matches the phonetic definition of being checked: the voicing is shut off abruptly by obstruent-like constriction.

3.5 Checked syllables and phonation in Vietnamese and Burmese

In Chapter 1, we have observed checked-like constituents in Vietnamese and Burmese. In both languages, the closed syllables associated with specific tones were analyzed as "checked syllables" by the studies of those languages. At the same time, there are also tones in those two languages that are associated with late-phased glottalization, but were not analyzed as "checked" in the grammar. Using the definition of phonologically checked constituents proposed in this dissertation, how will we analyze the closed syllables and late-phased glottalization in Vietnamese and Burmese? The current section addresses this question.

3.5.1 Vietnamese

Hanoi Vietnamese has eight tones. Vietnamese tones are derived from the same tonogenesis processes as Chinese (Haudricourt 1954; Michaud 2004). The tonogenesis processes and the correspondence between proto-Vietnamese and Modern Hanoi Vietnamese are in Table 3.16.

Table 3.16. Tonogenesis in Vietnamese

Stage 1	/*-∅/	/*-ʔ/	/-h/	/*-p, *-t, *-k/
Stage 2 (tonogenesis)	*ngang-huyền	*sắc-nặng	*hỏi-ngã	atonal /*-p, *-t, *-k/
Stage 3 (tone split)	t-: A1 (44 ngang) d-: A2 (32 huyền)	t-: B1 (24 sắc) d-: B2 (22 nặng)	t-: C1 (312 hỏi) d-: C2 (325 ngã)	t-: D1 (45 sắc) d-: D2 (21 nặng)

Syllables closed by /-p, -t, -k/, and tones D1 (45 sắc) and D2 (21 nặng) that occur in /p, t, k/-closed syllables fit the definition of checked syllables and checked tones proposed in the current chapter. The checked syllables and tones in Vietnamese have the same origin as the checked syllables and tones in Chinese languages.

Besides the checked syllables and tones, Vietnamese also has checked phonation. As discussed in Chapter 1 Section 1.3, Vietnamese has two tones that involve glottalization (Brunelle 2009). B2 (22 nặng) is short and has strong glottalization at the end of vowels. C2 (325 ngã) is long and has glottalization in the middle of vowels. What is the phonological status of the glottalization found in B2 (22) and C2 (325) tones? Is it a suprasegmental phonation feature realized with the tone (i.e. /V^{ʔ} 22/ and /V^{ʔ}V 325/), or is it a segment that is part of the syllable (i.e., /Vʔ 22/ and /VʔV 325/)?

The ʔ in VʔV with Tone C2 (325) is more likely to be suprasegmental. I have three pieces of evidence. They are not definitive, but taken together lean towards an analysis whereby the glottal stop in VʔV is suprasegmental.

1. If ʔ in VʔV is a segment, VʔV will be disyllabic, and the specific tone C2 (325) that VʔV bear has to be treated as two tones. However, Tone C2 (325) is originated from a single toneme < *hỏi-ngã (Table 3.16). It is the lower register of *hỏi-ngã after the tone split. Its counterpart, the upper register of < *hỏi-ngã – Tone C1 (312) – is undoubtedly a single toneme because it is borne by open syllables or syllables closed by nasals. Thus, it is quite unlikely that Tone C2 (325) can be split into two tones;

2. The glottalization in C2 (325) only occurs in Northern varieties of Vietnamese. In Southern varieties of Vietnamese, C1 and C2 tones are merged into one single tone, and the merged tone is not glottalized (Brunelle 2009; Kirby 2010). The analysis is more complicated if VʔV is disyllabic and the C2 tone contains two tones. It indicates that Southern Vietnamese changes all the disyllabic words (VʔV) into monosyllabic (V) words, and merged a single toneme C1 with two tonemes. Such an analysis is more complicated than

treating the glottal stop in VʔV as suprasegmental and having a rule that glottalization is lost in Southern Vietnamese;

3. The majority of words in Vietnamese are monosyllabic (Hwa-Froelich, Hodson, and Edwards 2002). If VʔV is disyllabic, we have to conclude that almost all the disyllabic words in Vietnamese are in the shape of V.ʔV (or Vʔ.V) and bear the Tone C2 (325). Although it is possible to still analyze VʔV as disyllabic and ʔ as segmental, it is not a favored analysis based on the overall phonological structure of Vietnamese.

The final-glottalization in tone B2 (22) is also likely to be suprasegmental, although the evidence is not definitive. B2 (22) tone occurs in both open syllables and syllables closed by nasals (VN̠). If the final glottalization in B2 (22) tone is a segment, we have to posit that VN syllables bearing B2 (22) tone have complex codas /ʔN/. However, Vietnamese otherwise does not allow complex codas (Hwa-Froelich, Hodson, and Edwards 2002). Such an analysis complicates the syllable structure in Vietnamese. However, note that this evidence is not strong enough to be conclusive. Evidence that shows syllables bearing B2 (22) tone behaving the same as other open syllables, but differently from VN syllables, will be stronger evidence to determine that the glottalization is suprasegmental in B2 tone.

If the glottalization in B2 (22) and C2 (325) tones is suprasegmental, we can say that B2 (22) has a checked phonation, and C2 (325) has a rearticulated phonation, because they are characterized by late- and middle-phased glottalization, respectively. Crucially though, the checked and rearticulated phonations do not contrast with each other independently of tone. They are associated with specific tones – B2 (22) and C2 (325), respectively. In summary, Vietnamese has both checked syllables/tones (-p, -t, -k closed syllables bearing D1 and D2 tones), and checked phonation (final-glottalization in B2 tone). I summarize the tone system and the checked constituents in Vietnamese in Table 3.17.

Table 3.17. Checked phonation, syllables, and tones in Vietnamese tones. The tonal values are from Kirby 2011.

Proto-Viet-Muong coda	Modern Vietnamese tones	Syllable structure	Checked?
∅	A1 (44 ngang)	V, VN	No
	A2 (32 huyền)	N	
*-ʔ	B1 (24 sắc)	V, VN	No
	B2 (22 nặng)	V, VN	Checked phonation
*-h	C1 (312 hỏi)	V, VN	No
	C2 (325 ngã)	V, VN	Rearticulated phonation
*-p, *-t, *-k	D1 (45 sắc)	Vp, t, k	Checked syllable Checked tone
	D2 (22 nặng)	Vp, t, k	Checked syllable Checked tone

The difference between the checked and the rearticulated phonation types in Vietnamese and those in Zapotec languages is the interaction pattern between phonation and tone. The checked and rearticulated phonations in Vietnamese covary with tone. The checked phonation only occurs in B2 tone, which has a falling contour, whereas the rearticulated phonation only occurs in C2 tone, which has a rising contour. In contrast, the checked and rearticulated phonations in Zapotec languages are contrastive independent of tone.

The difference between the checked syllable and checked phonation are in their phonological structure and historical origin. Checked syllables and tones refer to a restriction between syllable type and tone. Checked phonation refers to late-phased glottalization realized on vowels. Checked syllable is derived from /p, t, k/-closed syllables in Proto-Viet-Muong, whereas checked phonation is derived from /ʔ/-closed syllables.

3.5.2 Burmese

As discussed in 1.4, Burmese has four tones: high, low, "creaky", and "checked". Similar to Chinese languages and Vietnamese, there are three proto-tones in Proto-Burmese: *1, *2, *3 (Thurgood 1976). Syllables closed with /p, t, k/ are toneless. The "creaky" tone in Modern Burmese is derived from proto-tone *3, which has /ʔ/ as the syllable coda. The "checked" tone in Modern Burmese is borne by syllables closed by /ʔ/. The /ʔ/ is derived from the /*-p, *-t, *-k/ codas in Proto-Burmese.

Table 3.18. Tonogenesis in Burmese

Proto-tone	*1 (low)	*2 (high)	*3 /*-ʔ/	atonal /*-p, *-t, *-k/
Burmese	low	high	"creaky"-falling	"checked"-falling /ʔ/

As explained in Section 1.4, based on the vowel quality difference between syllables bearing "checked" tone and syllables bearing "creaky" tone, we can conclude that the glottalization in Burmese "checked" tone is segmental, whereas in "creaky" tone is suprasegmental. We can represent the "checked" tone as /V̂ʔ/ and the "creaky" tone as /V̂ˀ/.

Based on the phonological definition of checked in this dissertation, the "checked" tone and the /Vʔ/ syllable should be identified as a checked tone (with a falling F0 contour) and checked syllable. The "creaky" tone should be identified as checked phonation with a falling tone. There is not enough phonological evidence to say whether the falling tone borne by the checked syllable and the falling tone borne by the open syllable with checked phonation is the same tone. I summarize the correspondence between checked constituents and the four tones in Burmese in Table 3.19.

Table 3.19. Checked phonation, syllable, and tone in Burmese tones.

Proto-Burmese tone	Modern Burmese tones	Pitch contour	Checked?
*1 (low)	Low	Low	N
*2 (high)	High	High	N
3 (/-ʔ/)	“Creaky”	Falling	Checked phonation
atonal /*-p, *-t, *-k/	“Checked”	Falling	Checked syllable Checked tone

The “checked” tone and the /Vʔ/ syllable are qualified as checked tone and checked syllable because they form a phonotactic constraint between obstruent-closed syllable and specific tone. The phonation in “creaky” tone fits the definition of checked phonation in this dissertation, because vowels bearing a “creaky” tone has late-phased glottalization. Strictly speaking, the checked phonation borne by the “creaky” tone in Burmese does not contrast with other phonations independently of tone. The checked phonation is associated with a falling pitch contour, and is allowed in open and nasal-closed syllables. The two other tones that occur in the same phonological environment have modal phonation, but a low or a high pitch contour. The checked phonation thus does not contrast with modal phonation indpendently of tone in open or nasal-closed syllables. Checked syllables are associated with a falling pitch contour, just like syllables with checked phonation. However, syllables bearing checked phonation always differ from checked syllables in the segmental structure (i.e. $V^{ʔ}$ vs. Vʔ). Thus, there is is not a strict minimal pair of checked phonation and modal phonation with the tone being the same. Just as in Vietnamese, the checked phonation in Burmese is dependent on tone.

3.6 Phonetics of checked syllables and checked tones

This section reviews the phonetic properties of the checked syllables and tones in Chinese languages, which are identified based on the definition of checked proposed in Section 3.4. The current section will review four phonetic properties of checked syllables and checked tones in Chinese languages: F0, duration, voice quality, and vowel quality, as Section 2.4 did for checked vowels in Zapotec languages.

3.6.1 F0

First, in terms of F0, several studies have measured the F0 value for checked and unchecked tones. Comparing the F0 level and contour of checked and unchecked tones in different language varieties, we will see that checked tones can either have a distinct F0 contour from all non-checked tones, or overlap with some non-checked tones in the F0 space. I review languages that measure the F0 contour of checked and unchecked tones and summarize whether their F0 value overlap with each other.

Table 3.20. Tonal values of checked and unchecekeced tones in Chinese languages. All the examples assign the tone numerals (1-5) or pitch levels (L, M, H) based on F0 values[a].

Language	Checked syllable codas	Checked tones	Unchecked tones	Overlap or not
Nantong Jianghuai Mandarin	/k/	**HM**, LM [H]	**HM**, L [ML], H, LM [MLM], MH	Y
Cantonese Yue	/p, t, k/	**5(5)**, **3(3)**, **2(2)**	**55**, **33**, **22**, 21, 35, 13	Y
Zhangzhou Min	/p, t, k/	**41**, **22**, 221	**41**, **22**, 35, 51, 33	Y
South Taiyuan Jin	/ʔ/	**2(1)**, 42	**21**, 52, 53	Y
Taiwanese Min	/p, t, k/	**31**, 53	**31**, 55, 51, 24, 33	Y
Fuzhou Min	/ʔ/	**5(5)**, 24	**55**, 52, 32, 31, 342	Y
East Hefei Jianghuai Mandarin	/ʔ/	**5(5)**/4(4)/5(4)[b]	**55**/35, 31/212/21, 13/214/22, 52/51	Y
Taishanese Yue	/p, t, k/	45, 33, 42	45, 33, 42, 22, 31	Y
Shanghainese Wu[c]	/ʔ/	44 (T4), 24 (T5)	51 (T1), 33 (T2), 13 (T3)	N
Amoy Min	/p, t, k/	32, 4(4)	55, 35, 52, 31, 33	N
Meixian Hakka	/p, t, k/	32, 5(5)	44, 21, 31, 52	N
Yun'ao Min	/p, k, ʔ/	43, 45	44, 35, 52, 323, 33	N

[a] When a checked tone is extra short, studies usually only assign one Chao numeral value to it. I include the end value of those tones in parenthesis based on the acoustic description provided in the literature to indicate the contour shape of the tone.

[b] Slash means that the tone values vary by speakers.

[c] The Chao numerals of the tones are assigned based on the F0 track graph in Chen and Gussenhoven (2015).

Table 3.20 shows that it is common for one or all of the checked tones in a language to share the same F0 height and contour with the non-checked tones in the language. Even if the checked tones do not overlap with any of the non-checked tones in the F0 space, it is still common for checked tones to have a similar F0 height and contour with a non-checked tone (e.g. Amoy Min has checked tone 33 and non-checked tones 31 and 32). This is understandable given the tonal density of many Chinese languages: the more tones there are in the inventory, it is perhaps more likely for those tones to be phonetically more similar to one another.

3.6.2 Voice quality

In terms of voice quality, several studies reported that checked tones are realized with non-modal phonations. Differences in phonatory quality have been measured using Open Quotient (OQ) from electroglottography: (Pan 2017; Pan, Huang, and Lyu 2016; Pan and Lyu 2021; Shao 2012; Tang 2014), or using acoustic correlates of glottal constriction and voicing periodicity (e.g., spectral tilt, CPP, jitter, and shimmer) (Pan 2017; Pan, Huang, and Lyu 2016; Pan and Lyu 2021; Shao 2012). Shao (2012) measured OQ and jitter for checked vs. unchecked tones in Mexian Hakka and Fuzhou Min. In Meixian Hakka, the OQ of checked T32 has a larger OQ than unchecked tones by about 50% on average. In contrast, T5 has a lower OQ than all other tones at the end of the vowel, indicating that it is more constricted than other tones. Both checked T32 and T5 have higher jitter values than unchecked tones, indicating a noisier quality of checked tones in the language. Taking OQ and jitter data together, one can conclude that in Meixian Hakka, checked T32 is breathy, whereas checked T5 is creaky. In Fuzhou Min, checked T5 has lower OQ than other tones at the end of vowels. The voice quality of checked T24 is more variable: some words of checked T24 are produced with an OQ higher than unchecked tones, whereas others are produced with an OQ as low as checked T5 at the end of vowels. This indicates that Fuzhou Min checked T5 is glottalized, while checked T24 can be produced with either a breathy or creaky voice quality, depending on the word. In Eastern Hefei Jianghuai Mandarin, the OQ of checked T5(5) is lower than 0.5 for all speakers at the end of the vowel (Tang 2014). Checked T5(5) has on average the lowest OQ among all tones for older(but not for younger) speakers. This indicates that the checked tone in East Hefei Jianghuai Mandarin is produced with glottalization, at least among older speakers. In Zhangzhou Min, checked tones show evidence of creaky voice quality as determined by lower H1–H2 and irregular voicing, based on the visual inspection of waveforms and spectrums (Huang 2018).

When the coda of the checked syllables are prone to deletion, the voice quality of the vowels in the checked syllables are likely to become modal. For Taiwanese Min, Pan (2017) found that 80% of the glottal stop codas in checked syllables are deleted in both citation and sandhi forms. High-checked T5 is produced with less vocal fold contact than non-checked high-falling T51 in both citation (unexpectedly) and sandhi forms. As expected, low-checked T3 is produced with more vocal fold contact than non-checked low-falling T31 in citation form, and with less vocal fold contact than T31 in sandhi form. Pan and Lyu (2021) found that when coda of the checked syllables was deleted in the production, the H1–A3 values of the low checked tone (tone 5) increase, indicating a less restricted quality. This implies that Taiwanese Min checked tones are produced with a more modal-like voice when the coda is dropped in the production.

3.6.3 Duration

In terms of duration, checked tones tend to be shorter than unchecked tones. Checked tones have been found to be shorter than non-checked tones in Nantong (Ao 1993) and East Hefei Jianghuai Mandarin (Tang 2014); Zhangzhou (Huang 2018), Amoy (Lai 2016), Taiwanese (Pan 2017; Kuo 2013), and Fuzhou (Shao 2012) Min; Guangdong Cantonese (Zhu et al. 2008) and Taishanese (Cheng 1973; Tan 2016) Yue; Shanghainese Wu (Chen and Gussenhoven 2015); South Taiyuan Jin (Jia 2013); and Meixian Hakka (Shao 2012). Some languages have one checked tone and longer than the other: Shanghainese: T24 > T4(4) (Chen and Gussenhoven 2015); Amoy Min: T32 > T4(4) (Lai 2016); Meixian Hakka: T24 > T5(5); Fuzhou Min: T24 > T5(5) (Shao 2012); Yun'ao Min: T45 > T43 (Zhang 2017).

There also exist languages whose checked tones have similar or longer durations than unchecked tones. Zhangzhou Min has two checked tones: 41 and 221. In monosyllabic citation form, checked T221 is longer than the unchecked falling tones, T51 and 41. However, checked T41 is the shortest tone in the language. Likewise, Hong Kong Cantonese has both short and long checked tones (Bauer and Matthews 2017; Zhu et al. 2008). The long checked tone has a similar duration to the shortest unchecked tones in the language (269 vs. 284 ms), whereas the short checked tone is half of the duration of the shortest unchecked tone (123 vs. 284 ms) (Zhu et al. 2008).

3.6.4 Vowel quality

Huang (2018) found that, in Zhangzhou Min, high vowels acquire a diphthongal quality in checked syllables: /i/ becomes [iɛ] and /u/ becomes [uɤ]. In Hong Kong Cantonese, the long checked T33 and short checked T5 in Hong Kong Cantonese are conditioned by vowel quality and length: checked T5 has short vowels: /e, e^j,ɵ, ɐ, o, o^w/; checked T33 has long vowels: /iː, yː, ɛː, œː, aː, uː, ɔː/ (Bauer and Matthews 2017). In Nanjing Jianghuai Mandarin, compared with vowels in unchecked syllables, vowels in checked syllables have a higher F1 for /e, o, i, u, y/ and higher F2 for /e, o, u/ in older generation, and higher F3 for /o, u, y/ for all speakers (Yang and Chen 2018). They observe that the vowel quality difference is due to the glottal constriction gesture at the end of the vowel. The glottal constriction is accompanied by jaw lowering and consequently a lower and fronter tongue position, resulting in formant frequencies 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/.

3.7 Summary: the definitions of "checked"

What are the differences between checked syllables and tones in Chinese languages and checked vowels in Zapotec languages? First, in terms of phonological structure, "checked" in Zapotec languages refers to a phonation category. It can be reduced to a suprasegmental feature [+constricted glottis] that is phased late with respect to the vowel. Since phonation is usually realized on the nucleus of the syllable, i.e. the vowel, we say some Zapotec languages have checked vowels. In Chinese languages, "checked" is a descriptor of the phonotatic restriction between obstruent-closed syllable shape (checked syllables) and the specific tones of obstruent-closed syllables (checked tones). "Checked" can be reduced to the combination of two categories – closed syllables and tones.

Second, in terms of tones, all the Zapotec languages in the survey contrast checked phonation with another phonation (quasi-)independently of tone. In Chinese languages, determining whether this also holds depends on the analyst's choice of whether to consider the checked tones as allotones of unchecked tones. For languages such as Nantong Jianghuai Mandarin, where there is phonologically evidence that the checked tone (H) is an allotone of the unchecked tone (LM), checked syllables contrast with unchecked syllables (i.e. open syllables and nasal-closed syllables) independently of tone (e.g., /CVk LM/ vs. /CV LM/). For languages where we have phonological evidence that checked tones belong to phonemically different tones from unchecked tones, checked syllables do not contrast with unchecked syllables independently of tone. But, regardless of whether checked syllables in Chinese language contrast with unchecked syllables independently of tone, there is a fundamental difference between checked syllables and the checked phonation in terms of their relationship with tone: phonation and tone can be fully crossed in Zapotec languages, meaning that the checked phonation can bear the same set of tones as the unchecked phonations in the language. For Chinese languages, however, for a syllable to be qualified as checked, it never bears the exact same set of tones as unchecked syllables. The reason is that if checked syllables can bear any tone in the language, there is no need to posit a checked category in the language; we could simply analyze the language as a tonal language having open and closed syllables.

To summarize, I restate the phonetic characteristics of "being checked", and the phonological definition of "checked" in Zapotec languages, "checked" in Chinese languages, "checked phonation", "checked vowel", "checked syllable", and "checked tone" as following:

- Phonetic characteristics of being checked: An abrupt offset of voicing. The amplitude of the vowel decreases suddenly, usually reinforced by obstruent-like closures at the end of vowels;

- Checked phonation: A phonation type that is realized with glottalization at the end of vowels. It can be an allophonic realization of some kind of glottalized phonation in the language, or be contrastive with another glottalized phonation, which is usually rearticulated phonation with glottalization in the middle of the vowels;

- Checked vowel: A vowel that has checked phonation;

- Checked syllable: Syllables that are closed by obstruents and that are associated with distinctive tones from open syllables and sonorant-closed syllables, or with a subset of tones borne by open syllables and sonorant-closed syllables;

- Checked tone: Tones that are borne by checked syllables.

Chapters 1, 2, and 3 have addressed the first research question of this dissertation: on the phonological level, when is it necessary to posit a phonological constituent of "checked"? In the following chapters, the second research question is discussed: for the "checked" constituent that refers to a distinct phonological constituent, what are its phonetic properties and what acoustic cues do listeners use to perceive it? I approach the second question using original data from my field research on an under-studied Chinese language – Xiapu Min. Xiapu Min has checked syllables and checked tones. I study the phonetic properties of Xiapu Min checked syllables and tones in production and perception.

Chapter 4

Production of Xiapu Min checked tones in citation forms

4.1 Phonology of Xiapu Min

Xiapu (霞浦) Min is a variety of Eastern Min spoken in Xiapu County (Ningde, Fujian), China (see Figure 4.1 for the map retrieved from https://en.wikivoyage.org/wiki/File:Fujian_map.png and https://www.google.com/maps on 14 February 2022). There were 475,936 residents living in Xiapu County in 2021 (Xiapu County Bureau of Statistics 2021). Wen (2015) described the phonological system of Xiapu Min. As reported in Wen 2015, Xiapu Min has checked syllables that are closed by a glottal stop. There are two checked tones associated with checked syllables: T5(4) and T2(1). I 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. Xiapu Min has five unchecked tones: high-level T44, low-level T11, mid- and high-rising T23, 35, and falling T42. I measure the acoustic properties of Xiapu Min checked syllables tones to provide phonetic description to the checked syllables and tones defined in this dissertation. The data presented in the following chapters were collected by me and my collaborator, Shihong Ye, in Xiapu County from 2019 to 2021.

Figure 4.1. Xiapu County map.

The phonological report by Wen (2015) describes the Xiapu Min spoken in Songcheng District. The language consultants for my field research are from the downtown area of Xiapu Min. The phonological system of Xiapu Min based on the data from my field research differs from the one by Wen (2015). The differences are possibility due to the difference in the dialect variety and the generation of the speaker. Here, I present the consonant, vowel, and tone system of Xiapu Min.

There are 19 consonants in Xiapu Min. The consonant chart and the examples are in Tables 4.1 and 4.2. The glottal stop only occurs in the coda. There is allophonic glottal stop insertion to vowel-initial word in phrase-initial position.

Table 4.1. Consonant inventory of Xiapu Min

	Bilabial	Dental	Alveolar	Alveolo-palatal	Velar	Glottal
Stop	p p^h		t t^h		k k^h	ʔ
Fricative		θ		ɕ	x	
Affricate			ts ts^h	tɕ $tɕ^h$		
Nasal	m		n		ŋ	
Approximant	w		l			

Table 4.2. Examples of consonants in Xiapu Min

p	paʔ 5	百 "hundred"	θ	θu 35	数 "number"	m	ma 11	麻 "numb"
p^h	p^haʔ 5	打 "hit"	x	xu 35	富 "rich"	n	nan 11	南 "south"
t	te 42	底 "bottom"	ts	tse 42	紫 "purple"	ŋ	ŋa 11	牙 "teeth"
t^h	t^he 42	体 "body"	ts^h	ts^heʔ 2	蟹 "crab"	w	wa 42	我 "I"
k	ku 42	鼓 "drum"	ɕ	ɕjaʔ 5	那个 "that"	l	laʔ 2	六 "six"
k^h	k^hu 42	苦 "bitter"	tɕ	tɕjaʔ 5	这个 "this"			
ʔ	aʔ 5	鸭 "duck"	$tɕ^h$	$tɕ^h$ia 44	车 "car"			

In open syllables, there are 7 monophthongs, 8 diphthongs, and one triphthong in Xiapu Min (Table 4.3). Every monophthong also occurs in checked syllables. Monophthongs /y, e, u/ are not found in nasal finals. The distribution and examples of monophthongs, diphthongs, and triphthong in open, checked, and nasal-closed syllables are in Tables 4.4 and 4.5

Table 4.3. Vowel inventory of Xiapu Min

	Front		Back
High	i	y	u
Mid	e	ø	o
Low	a		

Table 4.4. Monophthong finals in Xiapu Min

i	kʰi 35	气 "air"	iʔ	tiʔ 2	直 "straight"	in	kin 44	金
						iŋ	iŋ 44	英 "English"
y	ky 35	锯 "chainsaw"	yʔ	tyʔ 5	柱 "pillar"			
e	ke 35	戒 "give up"	eʔ	eʔ 2	热 "hot"			
ø	kø 35	去 "go"	øʔ	øʔ 2	药 "drug"	øŋ	θøŋ 44	伤 "hurt"
a	ka 35	价 "price"	aʔ	taʔ 5	答 "answer"	an	an 44	安 "safe"
						aŋ	paŋ 35	半 "half"
u	ku 35	雇 "hire"	uʔ	tuʔ 5	督 "supervise"			
o	po 35	报 "report"	oʔ	toʔ 5	桌 "desk"	oŋ	poŋ 23	饭 "meal"

Table 4.5. Diphthong and triphthong finals in Xiapu Min

ia	θia 11	蛇 "snake"	iaʔ	θiaʔ 2	食 "eat"	iaŋ	tʰiaŋ 44	听 "listen"
iu	θiu 44	修 "fix"	iuʔ	niuʔ 2	肉 "meat"	iuŋ	kiuŋ 11	穷 "poor"
						ein	tein 23	电 "electric"
eo	keo 42	口 "mouth"						
ai	kai 35	盖 "cover"	aiʔ	θaiʔ 5	色 "color"	ain	kʰain 42	狗 "dog"
au	kau 35	到 "arrive"	auʔ	kauʔ 5	国 "country"			
ui	kui 23	跪 "kneel"						
ua	kua 44	瓜 "melon"	uaʔ	xuaʔ 5	法 "law"	uan	kuan 44	官 "officer"
uai	kuai 35	怪 "strange"						
oi	θoi 35	税 "tax"						

There are seven tones in Xiapu Min. I use the same tonal number by Wen 2015. I measure the F0 of the tones from multiple speakers to verify the tonal values. I establish the correspondence of the seven tones in Middle Chinese based on the reconstruction of Middle Chinese phonology by Zhengzhang (2003) and the reference dictionary by Institute of Linguistics (2018). The data from my field research agree with the reconstruction by Wen 2015 for the correspondence between Middle Chinese tone and Modern Xiapu Min tone. The examples of a minimal pair of the seven tones are in Table 4.6. Their F0 tracks produced by a female native speaker are drawn in Figure 4.2.

Table 4.6. Tones in Xiapu Min

Yīnpíng	44	θi 44	诗
Yángpíng	11	θi 11	时
Shǎng	42	θi 42	死
Yīngqù	35	θi 35	四
Yángqù	23	θi 23	是
Yīnrù	5	θi 5	湿
Yángrù	2	θi 2	实

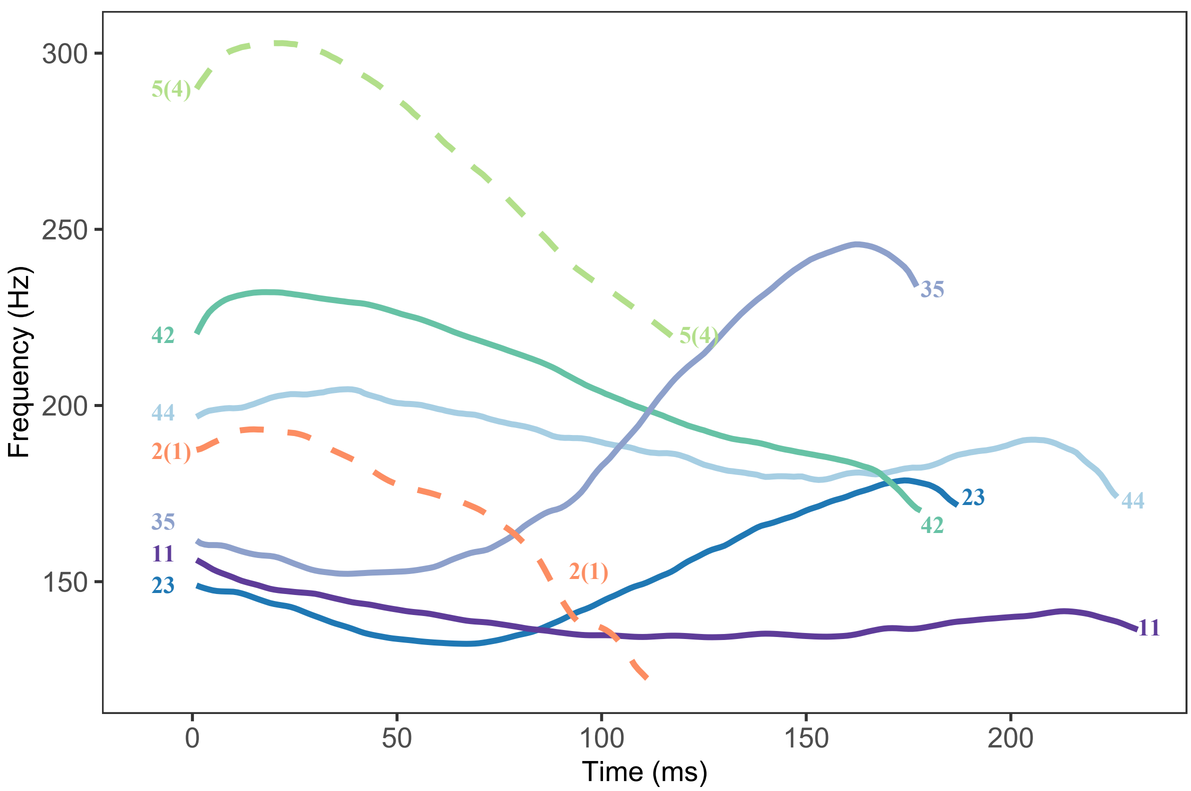


Figure 4.2. F0 track of /θi/ in seven tones by a female speaker.

4.2 The status of glottal stop

Glottal stop does not occur in the onset, but only occurs in the coda. There is glottal stop insertion before the vowel-initial word when the word is at phrase-initial position. Glottal stop is the only obstruent coda in Xiapu Min. The evidence that the glottal stop in Vʔ is a segment is from onset assimilation pattern, as has been found in other Min varieties (Section 3.2.1.2). In Xiapu Min, Vʔ syllables contrast with V and VN syllables in onset change pattern in disyllabic compounds. The phonological rules and examples are presented in Example (24)[1]. In a disyllabic compound, the /t/ onset of the second syllable becomes [ɾ] when it follows a V syllable (24a), [n] following a VN syllable (24b), and remains [t] when it follows a Vʔ syllable (24c). The /x/ onset of the second syllable is deleted when it follows a V syllable (24d), [ŋ] following a VN syllable (24e), and remains [x] when it follows a Vʔ syllable (24f). The fact that Vʔ syllables do not induce onset lenition or deletion in the second syllable, as the open syllables do, is evidence that ʔ is a segment. It functions as a consonantal barrier to prevent the onset lenition or deletion. The spectrograms for Example (24) are in Figure 4.3 and 4.4.

(24)	Rule	Example
(a)	/t/ → [ɾ] / CV$_{\sigma}$ + $_{\sigma}$___	/t^{h}e 42 tain 23/ → [t^{h}e 55 ɾain 23] 体重 “body weight”
(b)	/t/ → [n] / CVN$_{\sigma}$ + $_{\sigma}$___	/poŋ 44 toʔ 5/ → [poŋ 44 noʔ 5] 饭桌 “dining table”
(c)	/t/ → [t] / CVʔ$_{\sigma}$ + $_{\sigma}$___	/t^{h}eʔ 5 to 23/ → [t^{h}e 55 to 23] 铁路 “railroad”
(d)	/x/ → ∅ / CV$_{\sigma}$ + $_{\sigma}$___	/to 23 xe 35/ → [to 44 e 35] 路费 “traveling fee”
(e)	/x/ → [ŋ] / CVN$_{\sigma}$ + $_{\sigma}$___	/xein 44 xu 2/ → [xein 44 ŋu 5] 幸福 “happy”
(f)	/x/ → [x] / CVʔ$_{\sigma}$ + $_{\sigma}$___	/øʔ 5 xui 23/ → [øʔ 5 xui 23] 约会 “date”

1. Note that the phenomenon in Example (24) is undergoing change and loss. For example, Wen (2015) found that the onset change rules in Xiapu Min (Songcheng district) apply to high-frequency colloquial words, but not to the words used in a formal register.

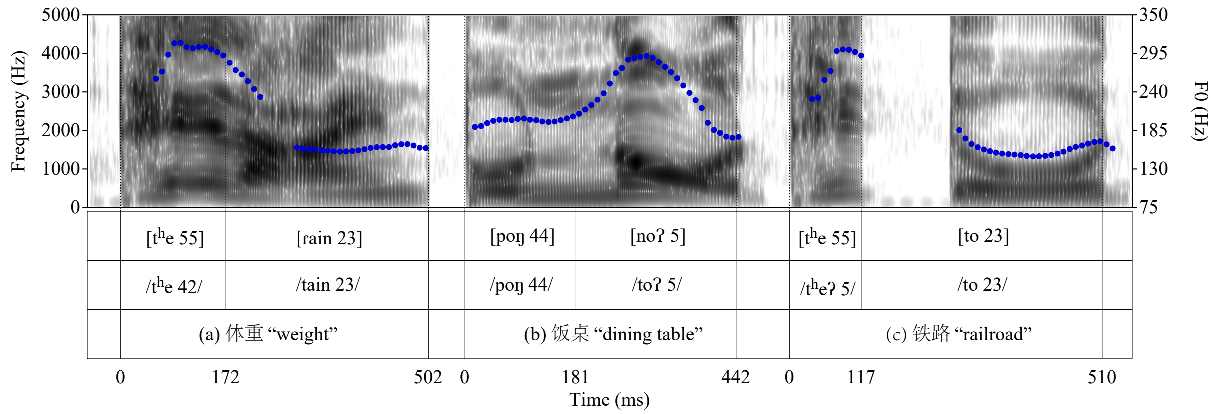


Figure 4.3. Spectrograms of /t/ onset assimilation for Examples (24a), ((24b), (24c).

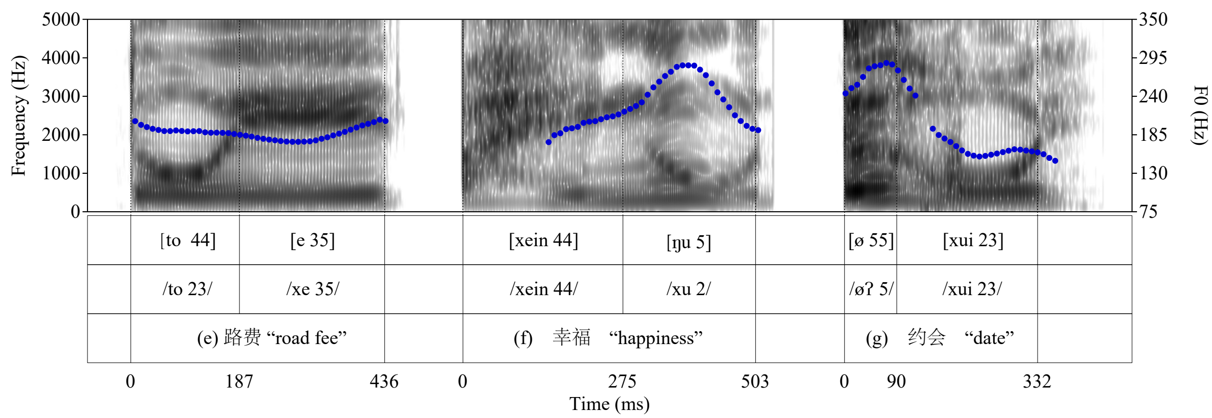


Figure 4.4. Spectrograms of /x/ onset assimilation for Examples (24d), (24e), (24f).

Given the evidence that the glottal stop in Vʔ is a segment, we can confirm that the Vʔ and tones associated with Vʔ fit the definition of checked syllables and tones proposed in Chapter 3. In the rest of this chapter, I present the acoustic analysis of Xiapu Min checked syllables and tones when they are produced in isolation in their citation forms. The results from the production study will answer: how do checked syllables differ from unchecked ones in terms of their F0 height and contour, phonation type, and duration?

4.3 Materials and Methods

4.3.1 Stimuli

Ten native speakers of Xiapu Min (five women, five men) with an average age of 53.5 (min: 46; max: 60) 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. All the experiments in this dissertation are approved by the Institutional Review Board of the University of California San Diego (protocol code: 190550; date of approval: 29 April 2019).

During the experiment, I asked the participants to produce minimal pairs of citation tones. There are 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. The complete list of stimuli is presented in Supplementary Material S1 at https://doi.org/10.17605/OSF.IO/3F5MN. A sample token of each word in the stimuli produced by the participants of the experiments is available in Supplementary Material S2 at https://doi.org/10.17605/OSF.IO/3F5MN. 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.

4.3.2 Acoustic Measures in Use and Criteria for Detecting Tracking Errors and Outliers

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

I 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). I 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). I predict that in Xiapu Min checked syllables, the glottal coda will trigger glottalization at the end of the vowel. Consequently, I 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 Vʔ token. We drew an F0 track for every Vʔ token and inspected whether there was pitch halving or doubling in the F0 track. When the pitch tracking failed, we excluded their F0 values from F0 analysis. These files are also excluded from H1*–H2* analysis because the correct estimation of H1*–H2* depends on 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 at https://doi.org/10.17605/OSF.IO/3F5MN. All the other supplementary tables for the current experiment can be found in Supplementary Material S3.

I 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, Jouan-Rimbaud, and Massart 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.

4.4 Results

Note that by my definition, 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, I will refer to the results for checked syllables and checked tones together as the results for checked tones. I chose "checked tone" as the reference because there are two checked tones that need to be differentiated.

The R script for data analysis and the raw data are available in Supplementary Materials S4 and S5 at https://doi.org/10.17605/OSF.IO/3F5MN. 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 (2.9% discarded), 125,700 (8.4% discarded), 125,700 (8.4% discarded), 119,838 (12.7% discarded), and 136,469 (0.6% discarded) 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.

4.4.1 F0

Figure 4.5 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.

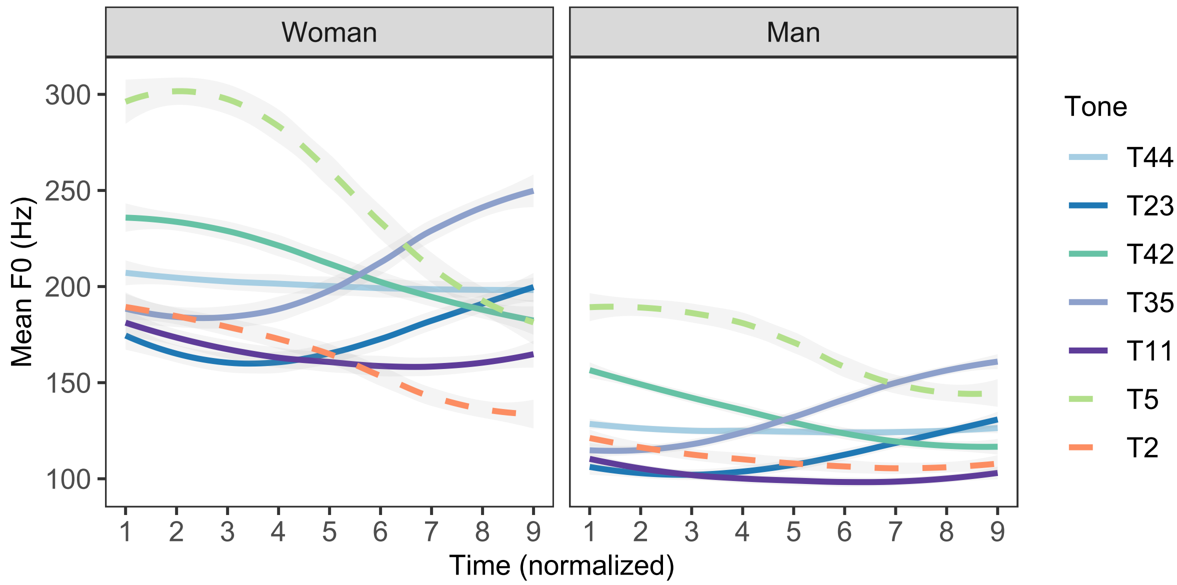


Figure 4.5. Average F0 track for five female and five male speakers of the seven tones in Xiapu Min. The contours are smoothed using the method "loess." The error ribbons represent 95% confidence interval.

Figure 4.5 shows that the F0 values in Hertz have large variation between females and males. We need to transform the F0 values in Hertz to a less varied scale in order to reduce between-speaker variation and establish a more uniformed representation of the tonal values. Log z-score has been found to be the most effective measure for reducing between-speaker variation among other F0 normalization methods (Zhu 2004), and it has been used in several studies (Duan and Jia 2015; Hu, Jia, and Liu 2012; Jia and Li 2012). The calculation of the log z-score was described in Section 4.3.2. 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 4.6a.

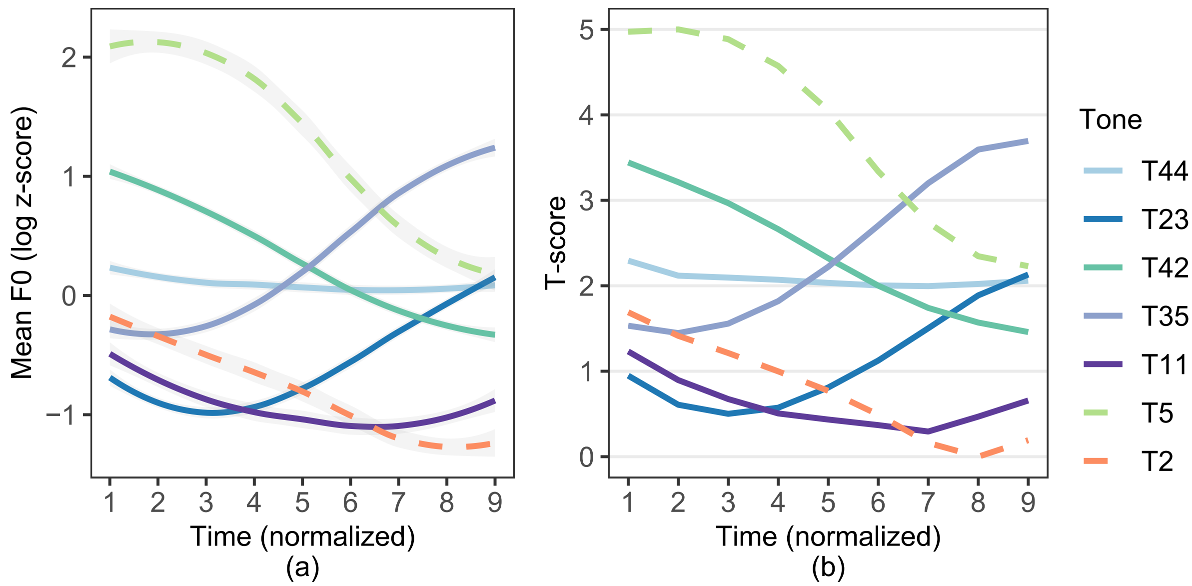


Figure 4.6. Log z-score (a) and T-score (b) of F0 over nine equally timed intervals of seven tones in Xiapu Min. The contours in (a) are smoothed using the method of "loess." The error ribbons represent 95% confidence interval.

Tonal values in Chinese languages are usually represented by Chao numerals. Shi, Ran, and Wang (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 (25). 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 4.7. 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.

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

Table 4.7. 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 (25) into Formula (26) 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 4.6a, 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 4.6b.

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

The T-score at Time Point 1 and 9 of each tone, and their corresponding Chao numeral based on Table 4.7, are listed in Table 4.8. Note that, based on the rules in Table 4.7, 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 4.8. 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)	T-Score	Chao numeral	Revised tonal values
T5	4.97	5	T53
	2.23	3	
T2	1.69	2	T21
	0.19	1	
T42	3.44	4	T42
	1.46	2	
T44	2.29	3	T33
	2.06	3	
T35	1.53	2	T24
	3.69	4	
T23	0.95	1	T13
	2.13	3	
T11	1.23	1[a]	T11
	0.66	1	

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

4.4.2 Voice quality

4.4.2.1 H1*–H2*

Figure 4.7 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 dissertation). The R code for the H1*–H2* model is in (27). Model (27) 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).

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

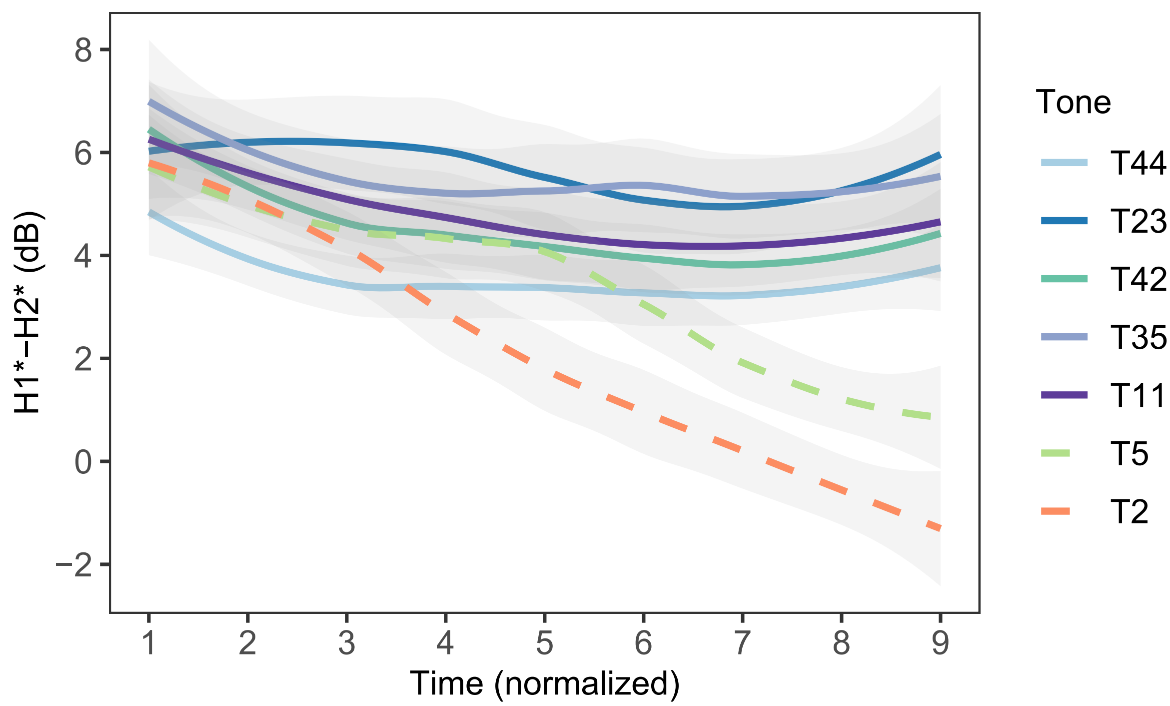


Figure 4.7. Average H1*–H2* track of the seven tones in Xiapu Min. The contours are smoothed using the method of "loess." The error ribbons represent 95% confidence interval.

The statistics of Model (27) 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.

4.4.2.2 HNR

Figure 4.8 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 (28). Model (28) 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).

(28) $lmer(HNR \sim Tone + (1|Participant))$

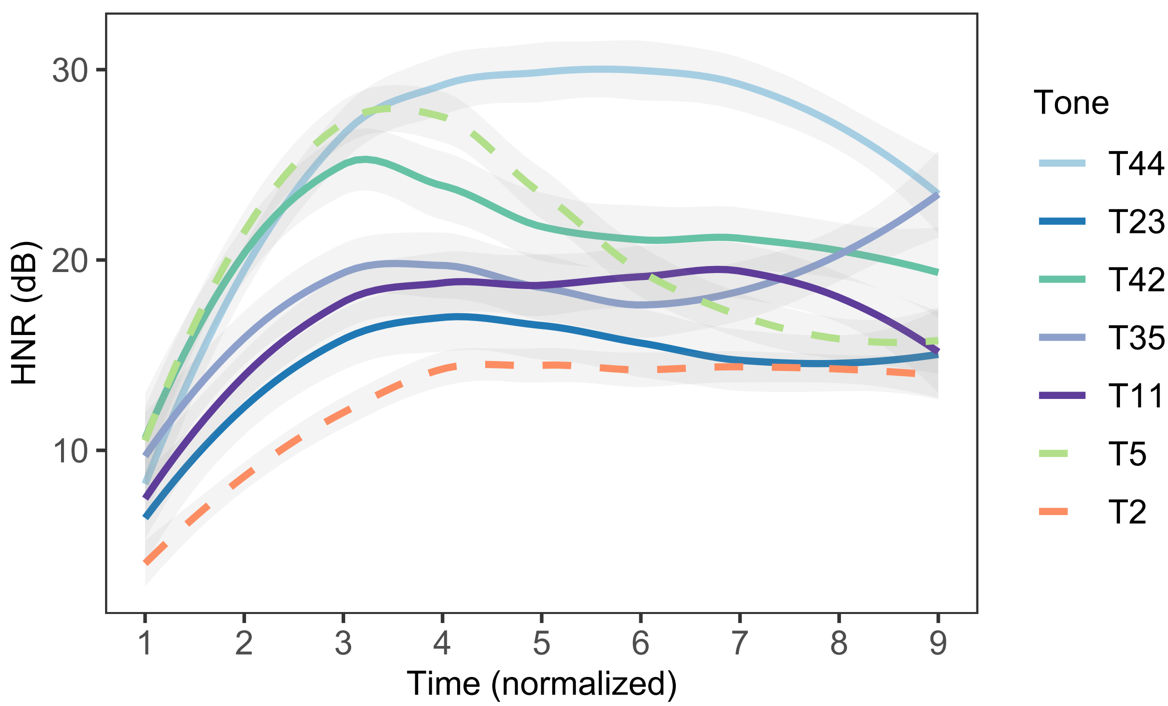


Figure 4.8. Average HNR track of the seven tones in Xiapu Min. The contours are smoothed using the method of "loess." The error ribbons represent 95% confidence interval.

The statistics of Model (28) are presented in Tables S10 and S11 in Supplementary Material S3. The results of Model (28) 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 4.8 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 4.9 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 4.9 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.

Table 4.9. 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 comprehensively. 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 4.9 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).

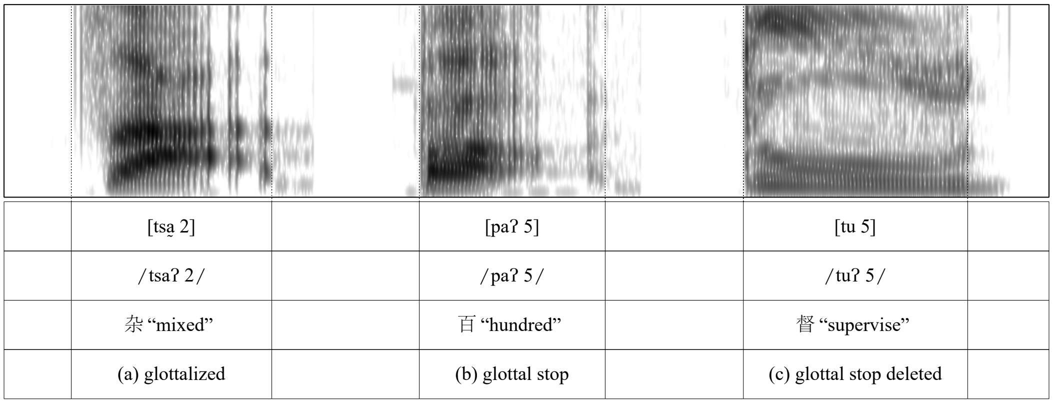


Figure 4.9. Spectrograms of glottal stops realized as (a) glottalized vowel, (b) full glottal stop and (c) deletion.

4.4.3 Duration

Figure 4.10 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 (29). No random intercept was included in the model because the random intercept did not improve the model and resulted in singular fit. Model (29) 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 (29) 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.

(29) $lm(Duration(z - score) \sim Tone)$

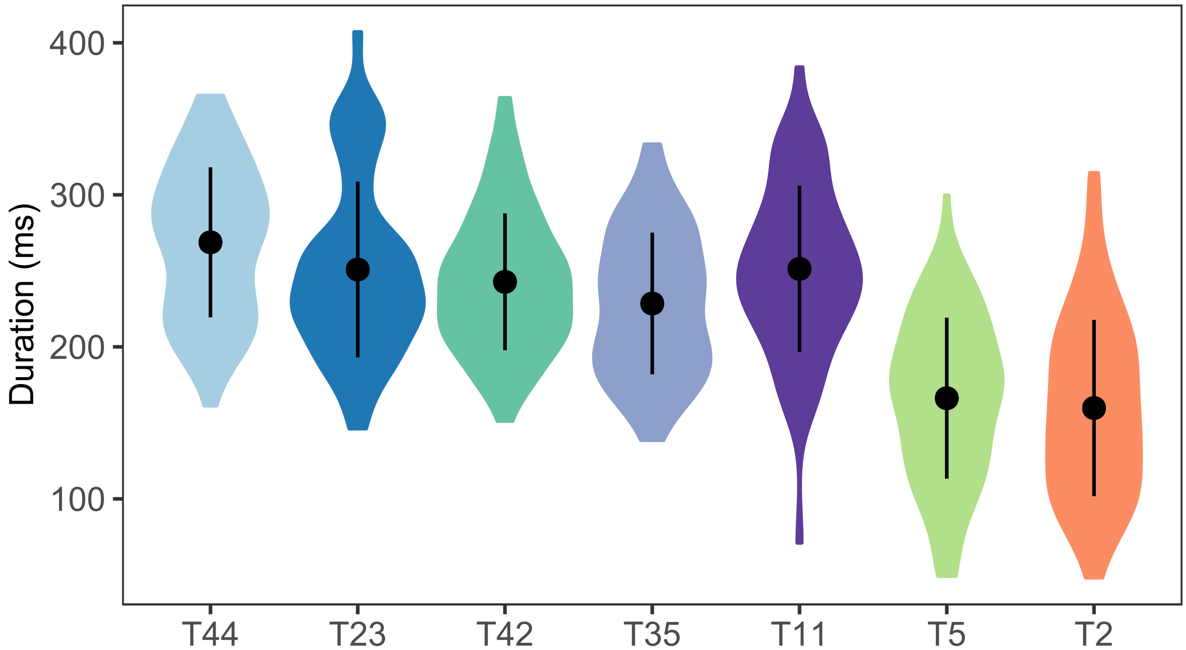


Figure 4.10. Average duration of the seven tones in Xiapu Min. The dots are the means. The error bars are the standard deviations.

4.4.4 Vowel quality

Figure 4.11 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 4.11 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 (30a) and (30a). The statistics of Model (30a) and (30a) 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.

(30) a. $lmer(F1 \sim Tone + Checkedness + (1|Participant))$

b. $lmer(F2 \sim Tone + Checkedness + (1|Participant))$

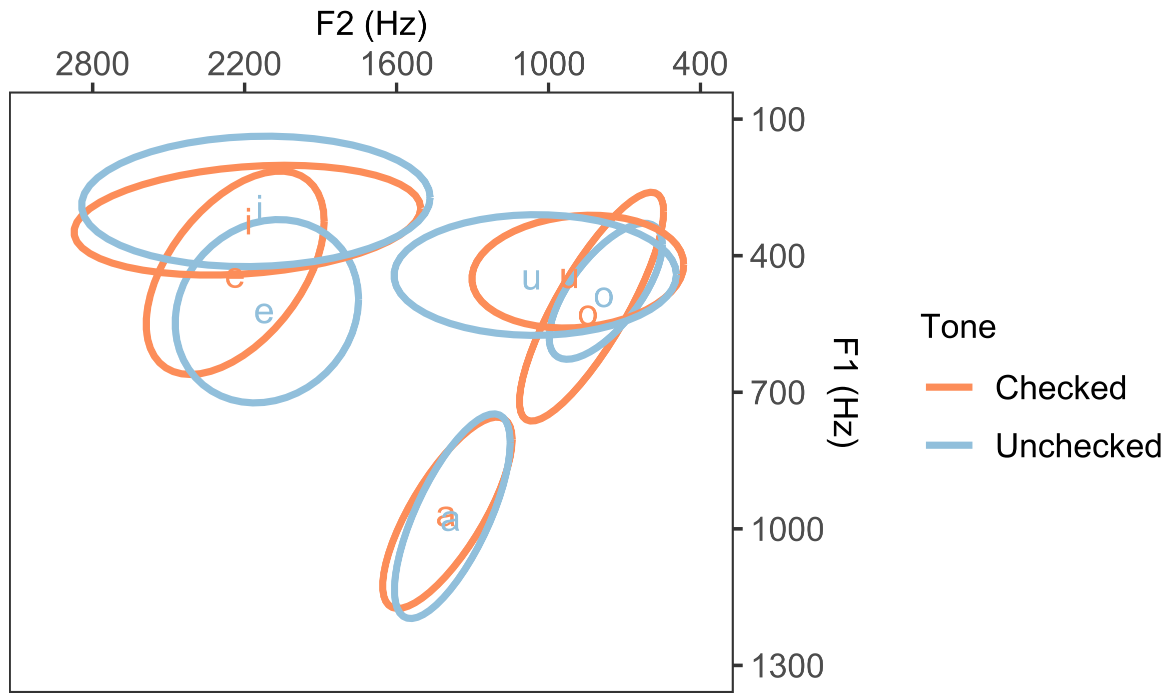


Figure 4.11. F1 and F2 distribution of vowels in checked and unchecked tones. The ellipses represent 95% confidence interval.

4.5 Summary

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.

Chapter 4, in part, is a reprint of the material as it appears in Chai, Yuan, and Shihong Ye. 2022. “Checked Syllables, Checked Tones, and Tone Sandhi in Xiapu Min” *Languages* 7, no. 1: 47. The dissertation author was the primary investigator and author of this paper.

Chapter 5

Production of Xiapu Min 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 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. Example (31) 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 (31a), whereas high-falling checked T5 becomes high-level unchecked T55 (31b) when they are followed by another tone in compounds.

(31) (a) /T2, T23, T44/ → [T44] / $\_\_\_\__{\sigma}$ + $_{\sigma}$X
(b) /T5, T35, T42/ → [T55] / $\_\_\_\__{\sigma}$ + $_{\sigma}$X
"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 me and my collaborator Shihong Ye.).

Example (31) shows that, in tone sandhi forms, checked T2 and unchecked T23 and T44 become phonologically neutralized as T44 (31a), while checked T5 and unchecked T35 and T42 become phonologically neutralized as T55 (31b). 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) or complete (e.g., Taiwanese Min: Chien and Jongman 2019). Given the large acoustic differences between checked and unchecked citation tones in Xiapu Min, I 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 (31a), 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 this chapter, my research question is: 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 question, I 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.

5.1 Materials and Methods

The same participants who produced the citation tones also participate in the production of the sandhi forms. The stimuli are 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. The complete list of stimuli is presented in Supplementary Material S1 at https://doi.org/10.17605/OSF.IO/3F5MN. A sample token of each word in the stimuli produced by the participants of the experiments is available in Supplementary Material S2 at https://doi.org/10.17605/OSF.IO/3F5MN. 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 the current experiment were covered in the production experiment in Chapter 4. During the experiment, I elicited the compound stimuli first and the one-syllable stimuli in the experiment in Section 4.3.1 next because I did not want to prime the participants with the underlying tone of the target syllables in the compound. The experiment presentation and equipment used are the same as in Section 4.3.1.

Table 5.1. Stimuli of neutralized sandhi tones in compounds for the production experiment (target syllable in bold).

Sandhi Rule	Contrast	Example	Gloss
/T2, T23, T44/ → [T44] / ___$_\sigma$ + $_\sigma$X	T2 vs. T23	/**xuʔ 2** tsɔŋ 44/ → [**xu 44** tsɔŋ 44]	服装 "clothes"
		/**xu 23** kain 44/ → [**xu 44** kain44]	护工 "caregiver"
	T23 vs. T44	/**to 23** kʰeu 42/ → [**to 44** kʰeu 42]	路口 "intersection"
		/**to 44** kʰeu 42/ → [**to 44** kʰeu 42]	刀口 "wound"
	T2 vs. T44	/**tsaʔ 2** ki 44/ → [**tsa 44** ki 44]	杂技 "acrobatics"
		/**tsa 44** kaŋ 44/ → [**tsa 44** kaŋ 44]	查岗 "check up"
/T5, T35, T42/ → [T55] / ___$_\sigma$ + $_\sigma$X	T5 vs. T35	/**θiʔ 5** tɕin 42/ → [**θi 55** tɕin 42]	湿疹 "eczema"
		/**θi 35** tɕia 42/ → [**θi 55** tɕia 42]	试纸 "test paper"
	T35 vs. T42	/**ka 35** kai 42/ → [**ka 55** kai 42]	价格 "price"
		/**ka 42** θe 42/ → [**ka 55** θe 42]	假设 "hypothesize"
	T5 vs. T42	/**kaʔ** 5 pain 42/ → [**ka 55** pain 42]	甲板 "deck"
		/**ka 42** tsein 42/ → [**ka 55** tsein 42]	假钱 "fake money"

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 4.3.2. After data exclusion, there were 78,960 (0.2% discarded), 74,299 (6.1% discarded), 74,299 (6.1% discarded), 71,638 (9.4% discarded), and 78,617 (0.6% discarded) 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 at https://doi.org/10.17605/OSF.IO/3F5MN. All the other supplementary tables for the current experiment can be found in Supplementary Material S3.

5.2 Results

5.2.1 Neutralization among T2, T44, and T23

The R script for data analysis and the raw data are available in Supplementary Materials S4 and S5 at https://doi.org/10.17605/OSF.IO/3F5MN. The first sandhi rule of Xiapu Min is /T2, T23, T44/ → [T44] / $\_\_\_\__{\sigma}$ + $_{\sigma}$X (31a). 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, Ripley, and Venables 2002).

The R code for the LDA models is in (32). 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.

(32) $lda(Tone \sim F0\_1 + F0\_2 + F0\_3 + H1^* - H2^*\_1 + H1^* - H2^*\_2 + H1^* - H2^*\_3 + HNR\_1 + HNR\_2 + HNR\_3 + Duration)$

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 5.1a 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 5.1b 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.

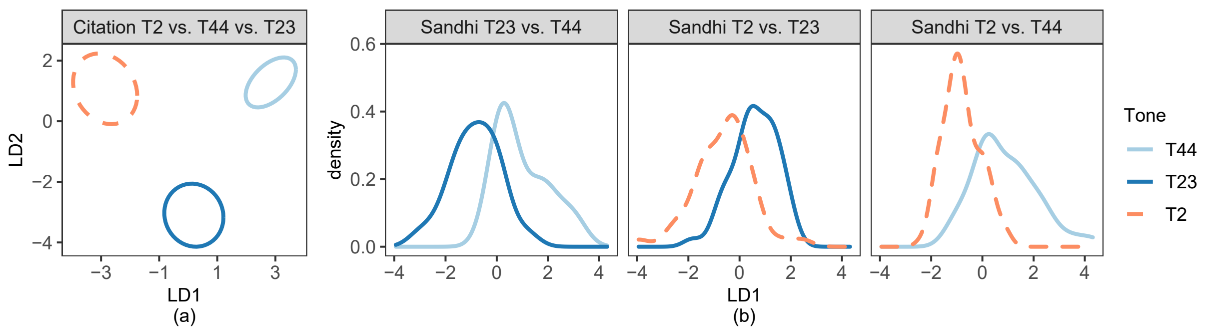


Figure 5.1. (a) is the first and second linear discriminant score (LD1 and LD2) distribution of T2 vs. T44 vs. T23 in citation forms. The ellipses represent 50% confidence intervals around the mean of each group. (b) is the LD1 distribution of T23 vs. T44, T2 vs. T23, and T2 vs. T44 in sandhi forms, respectively. The bandwidth of the density plots are calculated using the "rule-of-thumb" method by Silverman (1986). The smoothing kernel is "gaussian". 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 (32) and the correlations between the parameters and the linear discriminant scores are presented in Tables S17–S21 in Supplementary Material S3.

Figure 5.2 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.

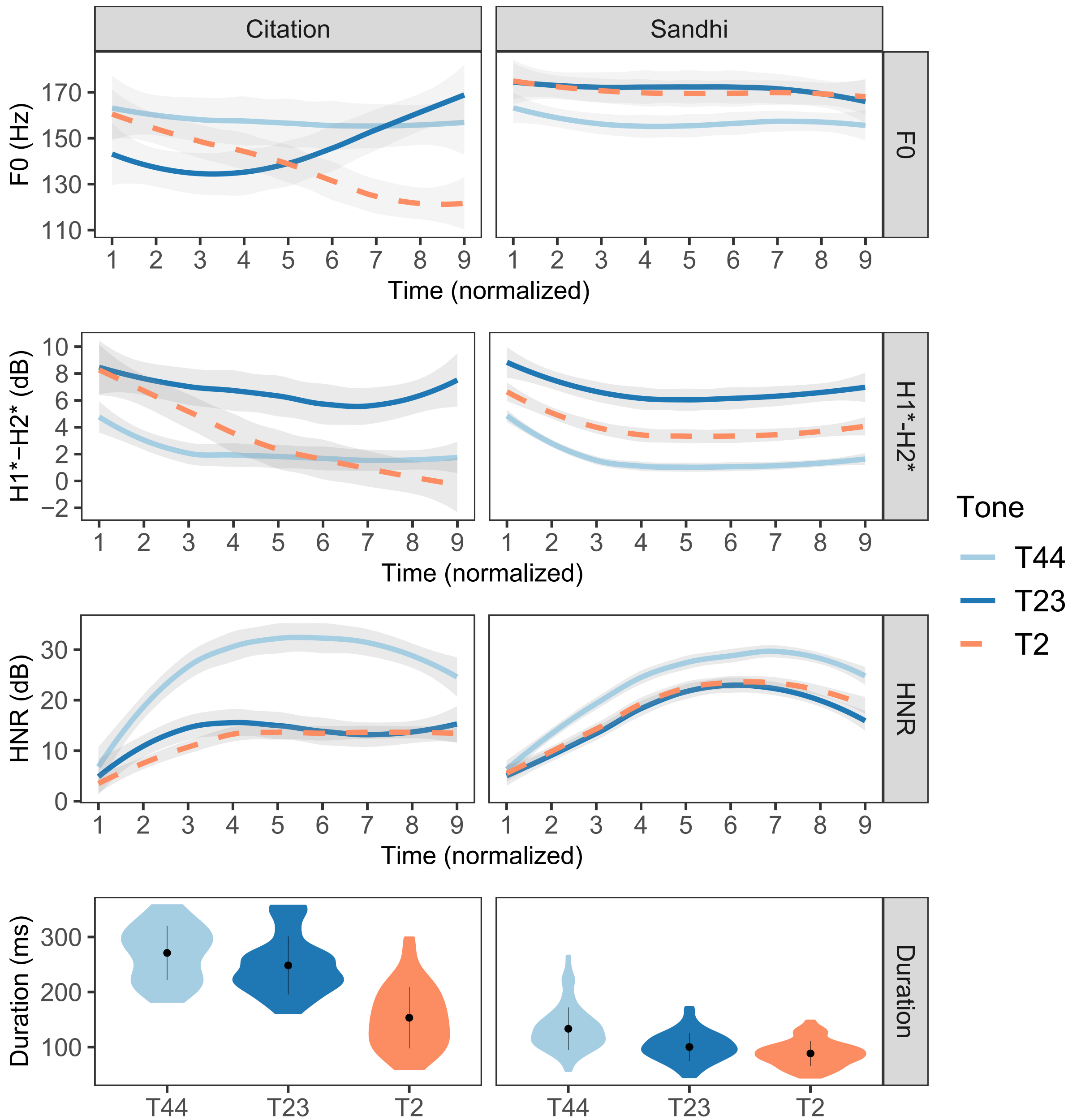


Figure 5.2. Acoustic parameter values of T44, T23, and T2 in citation and sandhi forms. The contours of F0, H1*–H2*, and HNR are smoothed using the method of "loess." The error ribbons represent 95% confidence interval. The dots and the error bars in the violin plots represent means and standard deviations.

5.2.2 Neutralization among T5, T42, and T35

The other sandhi rule of Xiapu Min is /T5, T35, T42/ $\rightarrow$ [T55] / $\_\_\__{\sigma} + {}_{\sigma}X$ (31b). 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 (32). 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 5.3a 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 5.3b 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.

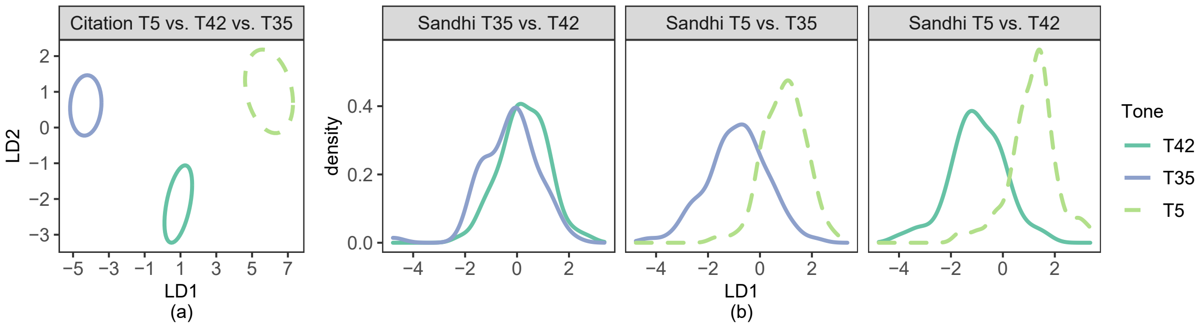


Figure 5.3. (a) is the LD1-LD2 distribution of T5 vs. T42 vs. T35 in citation forms. The ellipses represent 50% confidence intervals around the mean of each group. (b) is the LD1 distribution of T35 vs. T42, T5 vs. T35, and T5 vs. T42 in sandhi forms, respectively. The bandwidth of the density plots are calculated using the "rule-of-thumb" method by Silverman (1986). The smoothing kernel is "gaussian". 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 (32) and the correlation between all the parameters and the linear discriminant scores are presented in Tables S24–S28 in Supplementary Material S3.

Figure 5.4 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.

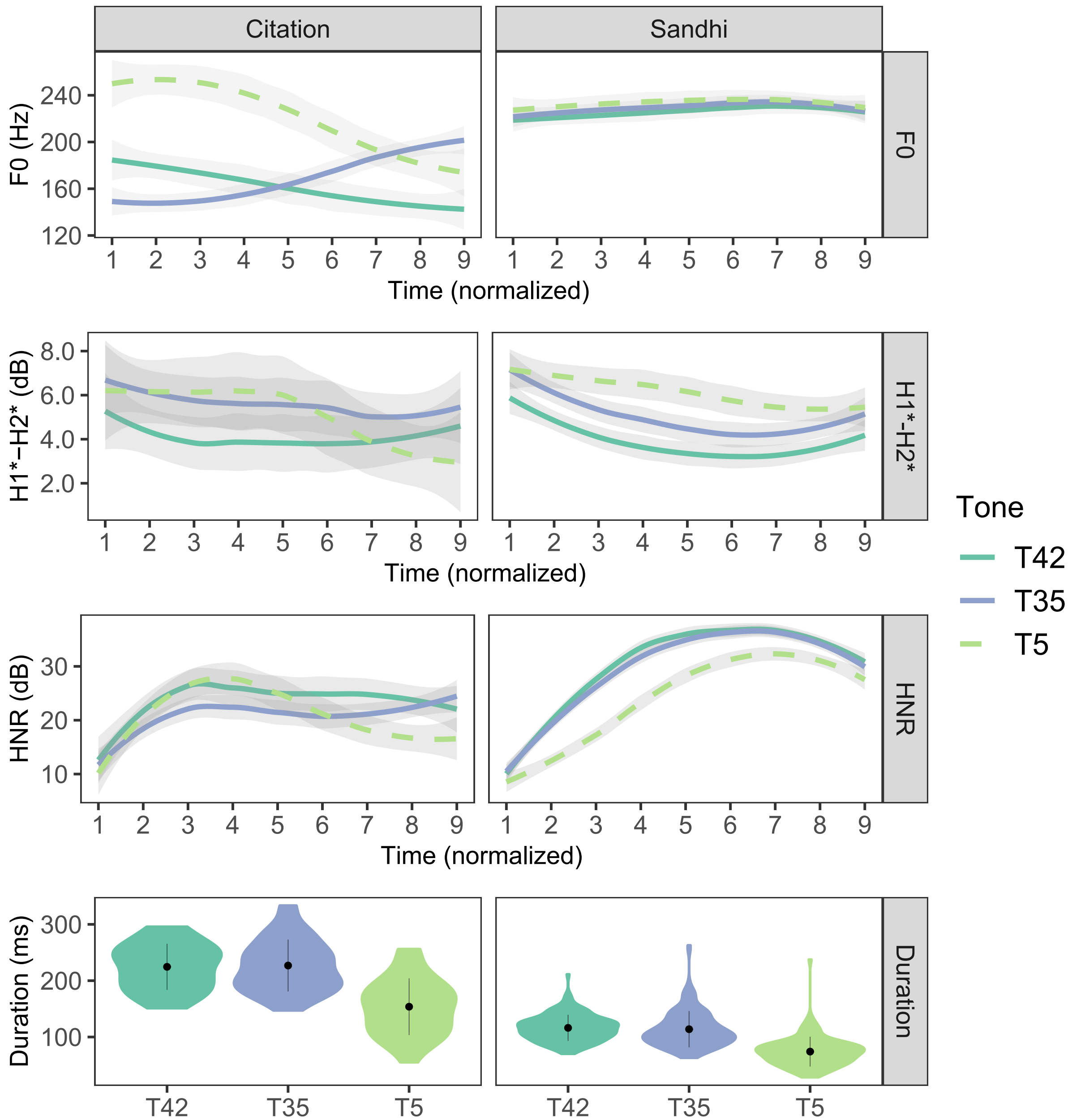


Figure 5.4. Acoustic parameter values of T42, T35, and T5 in citation and sandhi forms. The contours of F0, H1*–H2*, and HNR are smoothed using the method of "loess." The error ribbons represent 95% confidence interval. The dots and the error bars in the violin plots represent means and standard deviations.

5.3 Summary of the acoustic properties of checked tones in sandhi forms

Table 5.2 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 5.3 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%).

Table 5.2. 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, p Value)	Parameters
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)	
	T2 vs. T23	**68.69% (51.52%, <0.001)**	duration final F0 initial H1*–H2*
	T2 vs. T44	**76.61% (52.63%, <0.001)**	duration initial and final HNR
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)	
	T5 vs. T35	**80.35% (51.45%, <0.001)**	duration initial and mid HNR
	T5 vs. T42	**86.79% (53.77%, <0.001)**	duration initial and mid HNR

Table 5.3. 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	Ratio of Checked:Unchecked
Citation	166	160	269	243	228	251	251	66%
Sandhi	74	89	133	116	114	100	NA	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 5.2). 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 5.2). 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 (Figure 5, Figure 11 and Figure 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 5.2). 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.

5.4 Summary of the production experiments

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 CVʔ syllables. Cross-linguistically, vowels tend to be shorter in closed syllables than in open syllables (Farnetani and Kori 1986; Rietveld and Frauenfelder 1987; 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 CVʔ 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. Example (24) in Section 4.2 showed that CVʔ behaves differently from open syllables in phonological transformations, supporting the idea that the glottal stop in CVʔ 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. As discussed in Chapter 3 Section 3.4, across Chinese languages, there is a tendency for checked syllables to become unchecked. As proposed by Zhu et al. (2008) and shown in Figure 3.1, there are three possible paths for the checked syllable sound change: Type I: Checked syllable lengthening → Oral coda lenition to glottal coda → Glottal coda loss; Type II: Oral coda lenition to glottal coda → Glottal coda loss → Checked syllable lengthening; Type III: Oral coda lenition to glottal coda → Checked syllable lengthening → 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: the glottal stop is lost first, then syllable lengthening takes place.

To conclude: in Chapters 4 and 5, I provide 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 the following chapters, I will present the results of the perception studies that I conducted, in which I manipulate F0, phonatory quality, and duration separately in sound signals for both citation and sandhi checked tones. The perception studies allow us to directly test whether short duration, glottalization, and F0 are independent cues of checked tones in Xiapu Min.

Chapter 5, in part, is a reprint of the material as it appears in Chai, Yuan, and Shihong Ye. 2022. “Checked Syllables, Checked Tones, and Tone Sandhi in Xiapu Min” *Languages* 7, no. 1: 47. The dissertation author was the primary investigator and author of this paper.

Chapter 6

Perception of Xiapu Min checked tones in citation forms

6.1 Introduction

Knowing the acoustic properties of Xiapu Min checked tones in the production of their citation forms, I explore the phonetics of Xiapu Min citation checked tones from the perspective of perception in the current chapter. In Chapter 4, we see that Xiapu Min citation checked tones differ from unchecked tones in terms of F0, duration, and voice quality. Low-checked T2 has low-falling F0 contour and high-checked T5 has high-falling F0 contour, and these contours do not overlap with unchecked tones in the F0 space. Checked tones are also shorter and more glottalized than unchecked tones. In the current chapter, I ask: do the listeners of Xiapu Min make use of all these three acoustic cues when identifying checked tones? And if so, do they rely on one cue more over the others?

6.1.1 Studies on perception of checked tones

Few studies have tested the acoustic cues that listeners use for identifying *checked* syllables or tones. Here I review studies that have tested the perception of checked tones or tones that have similar phonetic properties to Xiapu Min checked tones. The results of these studies could be comparable to the results of Xiapu Min checked tone perception. Thus, I review studies that explore the perceptual cues for tones that are borne by syllables that are closed by an oral or a glottal stop, or by syllables with vowels ending in glottalization. Such tones include the checked tone in Burmese (Gruber 2011), the B2 (22) nặng tone in Vietnamese (Brunelle 2009), the creaky tone in White Hmong (Garellek et al. 2013), and the glottalized tone in Sgaw Karen (Brunelle and Finkeldey 2011).

Gruber (2011) conducted production and perception experiment on Burmese checked tone. As discussed in Section 1.4, Burmese has four tones: High, Low, "Creaky", and "Checked". As discussed in Section 3.5.1, based on the criteria of checked phonation and syllable/tone in this dissertation, the "creaky" tone in Burmese can be reanalyzed as having a "checked" phonation (/V^{ʔ}/). The "checked" tone indeed fits the definition of checked tone, and occurs in checked syllables (/Vʔ/). Leaving the phonological analysis of "Creaky" and "Checked" tones aside, "Creaky" and "Checked" tones have similar phonetic properties as Xiapu Min checked tones. They both have a falling F0 contour, which is different from the High and Low tones. They are also both glottalized at the end of the vowel. The duration of the four tones ranks as High & Low > "Creaky" > "Checked" in isolation and phrase-final positions. "Checked" tone and "creaky" tone occur in syllables with different vowel qualities. Syllables bearing "checked" tones can have centralized vowels and diphthongs whereas syllables with other tones only have peripheral monophthongs. Gruber's (2011) perception study resynthesized F0 height and contour, and duration based on Low (modal), High (breathy), and "Creaky"-toned tokens. The results show that phonation distinguishes the "Creaky" tone from the High tone. Average F0 distinguishes the High tone from the Low tone. However, "creaky" and "checked" tones are indistinguishable using the stimuli attested. "checked" tone responses were elicited when the stimuli were short and creaky, but there was not a condition that predominantly elicited "checked" tones. The possible reasons why Gruber (2011) did not find the "checked" tone being distinguished from the "creaky" tone in perception is that the stimuli did not simulate the acoustic differences between the "checked" tone and the "creaky" tone found in the production. "Checked" tone differs from "creaky" tone in two aspects: duration and vowel quality. However, neither of these were manipulated systematically in the stimuli. The shortest condition in the stimuli is based on the average duration of the "creaky" tone produced by a single speaker (175 ms). For the same speaker, their average duration for checked tone is 121 ms (calculated based on the raw data provided in Gruber 2011). And in terms of the vowel quality, the vowel quality in syllables bearing "checked" tone is more centralized compared with the "unchecked" (high, low, and "creaky") tones. However, the base of the stimuli are tokens with the three "unchecked" tones. The stimuli might have biased the listeners to "unchecked" tone responses because the vowel quality resembles the vowel quality of "unchecked" tones more. If future studies include shorter duration conditions and include "checked" tone tokens in the base stimuli, more "checked" tone responses might be elicited and distinguished from the "creaky" tone responses.

Vietnamese has three tones that are phonetically similar to the checked tones in Xiapu Min – B2 (22 nặng), D1 (45 sắc), and D2 (21 nặng). Tone B2 (22) has full glottal stop closure or strong glottalization at the end of the vowel. Tone D1 (45) and D2 (21) are borne by syllables closed by /p, t, k/. As discussed in Section 3.5.1, the B2 tone can be analyzed as having a checked phonation, and its checked phonation is in contrast with the rearticulated phonation in Tone C2 (325 ngã). Tones D1 (45) and D2 (21) fit the definition of checked tone and occur in checked syllables. All three tones have shorter duration than the other tones in the language. Each tone also has different F0 height and contour from other tones in the language (Michaud 2004; Brunelle 2009). Note that B2 (22) has strong glottalization in the end and C2 (325) has strong glottalization in the middle only in the Northern variety of Vietnamese; in Southern Vietnamese, these tones are not glottalized. Brunelle (2009) tested how phonation (modal, mid-glottalization, final-glottalization), and F0 onset, offset, and contour shape affect Northern and Southern Vietnamese listeners' perception of tones. The results show that for both groups, final-glottalization predominantly elicited the B2 (22) tone responses, regardless of the F0 condition. For Northern Vietnamese listeners, a falling or level F0 contour with mid-glottalization also elicits B2 (22) tone responses, though not as many as stimuli with final-glottalization. The results from Brunelle (2009) indicate that final-glottalization is a more important cue than F0 for identifying the low glottalized B2 (22) tone perception by both Northern and Sourthern Vietnamese listeners.

The creaky tone in White Hmong is phonetically comparable to the checked tones in Xiapu Min. White Hmong creaky tone is distinguished from other tones in the language in the F0 space. It is relatively shorter than other tones, and is glottalized at the end of the vowel (Huffman 1987; Ratliff 2010; Esposito 2012; Garellek et al. 2013). However, later work by Garellek and Esposito (2021) reported that the creaky tone is the shortest tone only on vowel /e/. It becomes the second longest tone when it is on vowel /u/. In addition, the creaky tone is indistinguishable from the low (modal) tone in H1*–H2* and CPP, indicating that creaky tone is not characterized by a glottalized quality. Garellek et al. (2013) tested how F0, phonation (modal, breathy, creaky), and duration affect listeners' perception of creaky tones in White Hmong. Garellek et al. (2013) resynthesized the F0 of naturally produced creaky and low modal tones so that their F0 is confusable with each other. Each token was also resynthesized into short and long durations. The results of the tone identification task show that the presence of creak did not have a significant effect on the identification of the creaky tone. Tokens with a low F0 in the end, a falling F0 contour, or a shorter duration elicited significantly more creaky tone responses. The results by Garellek et al. (2013) illustrate that listeners of White Hmong rely on the F0 and the duration cues more than glottalization for creaky tone identification. This is in accordance with the production results found in Garellek and Esposito 2021, such that glottalization is not a stable characteristics of the creaky tone in production either. Combining the creaky tone perception and production results together, it might also be evidence for a sound change in progress. Glottalization used to be present in the production of creaky tone, as indicated by previous studies (Huffman 1987; Ratliff 2010; Esposito 2012). However, the listeners are no longer using glottalization as a cue when perceiving creaky tone. And at present day, glottalization is not consistently produced in production either.

Sgaw Karen has six tones with four phonations: high-rising modal T1, mid-falling modal T5, high-falling breathy T2, high-falling breathy T6, mid-falling creaky T3, and mid-falling glottalized T4 (Brunelle and Finkeldey 2011). The mid-falling modal (T5), creaky (T3), and glottalized (T4) tones have similar F0 onsets, but differ in F0 offset; so do the two high-falling breathy tones (T2 and T6). The glottalized tone is phonetically similar to the checked tones in Xiapu Min: it has a more constricted quality than other tones in general (represented by lower H1*–H2* values), and is characterized by strong glottalization in the end. It also has the shortest duration among the six tones. The creaky tone, on the other hand, has mid-glottalization. Brunelle (2009) lengthened the glottalized tone and shortened the creaky and high-rising modal tones, so that the duration of modal, glottalized, and creaky tones became confusable. Then, they superimposed twenty-five different F0 contours varying in F0 onset and offset on the base tokens and the duration-manipulated tokens . Brunelle (2009) analyzed the results from all the duration and F0 manipulations using Classification Tree. The selected decision tree shows that tokens with mid- and final-glottalization, and short duration lead to glottalized tone decisions. Tokens with mid- and final-glottalization, long duration, and low F0 offset lead to creaky tone decisions. The results indicate that listeners use glottalization and short duration as the cue for glottalized tone in Sgaw Karen. Whether the glottalization is in the middle or the end of the vowel does not influence the perception. F0 is not a relevant cue for glottalized tone identification. Creaky tone is differentiated from glottalized tone primarily by its longer duration. The results from Brunelle (2009) illustrate that cue that is present in production can be redundant in perception – although a difference in the timing of glottalization is present in the production of creaky and glottalized tone, it is not used when listeners are differentiating these two tones in perception. Although the glottalized tone and the creaky tone differ (slightly) in slope and F0 offset, F0 is not a relevant cue for glottalized tone identification.

The above studies demonstrate that F0 (in White Hmong, Vietnamese), duration (in Burmese, White Hmong, Sgaw Karen), and phonation (in Burmese, Vietnamese, Sgaw Karen) all can be used for the identification of checked-like tones. However, these three cues are not necessarily used simultaneously. Even when two cues are used simultaneously, their importance might differ (as in Vietnamese). In order to determine how F0, duration, and phonation influence listeners' perception of checked tones in citation forms in Xiapu Min, I conduct a tone identification experiment with these three parameters manipulated.

6.2 Method

6.2.1 Stimuli

The stimuli used for the current experiment are available in Supplementary Material S6 at https://doi.org/10.17605/OSF.IO/3F5MN. In order to test the independent effect of F0, phonation, and duration on checked tone identification, I resynthesized stimuli varying in F0, phonation, and duration orthogonally. There are five F0 conditions: low-falling (T2-like), high-falling (T5-like), low-level (T11-like), mid-level (T44-like), and mid-falling (T42-like); two duration conditions: long and short; and two voice quality conditions: modal and glottalized. The base stimuli is a token of /θi 44 (mid-level)/ "诗 (poem)" produced a female speaker (Speaker #3 in the production study in Sections 4 and 5). The F0 contour and duration values are based on another female speaker (Speaker #1 in the production study in Chapters 4 and 5), who is judged as having a prototypical production of the seven tones by a native speaker of Xiapu Min. Her production was not used as the base token for resynthesis because of its poorer recording quality.

The five F0 conditions are based off tones that overlap with the checked tones in the F0 space. T2 (low-falling) and T5 (high-falling) are checked tones themselves. T11 (low-level) have similar F0 height as T2. T42 (mid-falling) has a similar F0 contour to T2 and T5. The F0 height of T44 (mid-level) is between those of T2 and T5. The F0 of T11, 42, and 44 can be confusable with checked tones T2 and 5 in perception. Creating F0 contours of T11, 42, and 44 in addition to T2 and 5 can test when the F0 of the test stimuli is similar, but not identical to the checked tones, whether the responses to the checked tones will decrease. The tonal values are based on Speaker 2's production of [θi] in those five tones. The F0 values are in Table 6.1.

In terms of the duration conditions, the short condition has a vowel 120 ms. The long condition has 240 ms. These values are based off the duration of the shortest (T2) and the longest

(T11) tones among the seven tones realized on /θi/ produced by Speaker #1.

In terms of the two voice quality conditions, the base token has modal phonation. The glottalized phonation was created by lowering and jittering the second half of the vowel. The F0 values and corresponding time point of each pitch point are in Table 6.2. Glottalization was added to the second half of the vowel because, in the citation forms of the checked tones produced by the nine speakers in the production experiment (Chapter 4), the average proportion of glottalization is 50% of the vowel. The glottalization is usually found at the end of the vowel in the production. I simulated the glottalization by lowering and jittering the F0, because prototypical creaky voice is produced with low and irregular F0, as well as a constricted glottis (Keating, Garellek, and Kreiman 2015). Previous studies have also used this method to simulate creaky voice, and have proved its effectiveness in eliciting a percept of creaky voice. For instance, Frazier (2009) simulated the glottalization in glottalized vowels [V́V̰] in Yucatec Maya by inserting a pitch point with an extra-low F0 value (35 Hz) in the middle of the vowel. The stimuli with such kind of glottalization did elicit more responses for glottalized vowels from the listeners, which proves the effectiveness of using low F0 to elicit a percept of glottalization. Huang (2020) tested the independent effect of extra-low F0 and irregular F0 on Mandarin tone identification, and found that extra-low F0 can facilitate the identification of the creaky tone (T3) and hinders the identification of the modal tones (T1, 2, 4). Irregular F0 hinders the identification of modal tones. The results from Huang 2020 suggest that extra-low F0 and irregular F0 can elicit creak percept independently. The intended F0 contours in Tables 6.1 and 6.2 of all the conditions in the stimuli are schematized in Figure 6.1.

Table 6.1. F0 of the modal tokens in the stimuli. The F0 values are assigned to the relative time point in the vowel. P1 to P9 refer to the first to the ninth F0 point. P1 to P9 are at the 1/10 to 9/10 of the duration of the vowel.

		Low-falling (T2-like)	High-falling (T5-like)	Low-level (T11-like)	Mid-falling (T42-like)	Mid-level (T44-like)
F0 point	Time point					
P1	1/10	167	294	150	230	204
P2	2/10	165	303	144	231	203
P3	3/10	163	299	140	226	200
P4	4/10	160	289	136	218	198
P5	5/10	158	275	134	208	197
P6	6/10	157	261	135	198	196
P7	7/10	155	244	136	191	195
P8	8/10	150	231	139	186	194
P9	9/10	141	219	140	173	193

Table 6.2. F0 of the glottalized tokens in the stimuli. The F0 values are assigned to the relative time point in the vowel. P1 to P12 refers to the first to the twelfth F0 point. P1 to P5 are at the 1/10 to the 5/10 of the duration of the vowel. P6 to P12 are at the 9/16 to the 15/16 of the duration of the vowel.

		Low-falling (T2-like)	High-falling (T5-like)	Low-level (T11-like)	Mid-falling (T42-like)	Mid-level (T44-like)
F0 point	Time point					
P1	1/10	167	294	150	230	204
P2	2/10	165	303	144	231	203
P3	3/10	163	299	140	226	200
P4	4/10	160	289	136	218	198
P5	5/10	158	275	134	208	197
P6	9/16	135	135	135	135	135
P7	10/16	95	95	95	95	95
P8	11/16	110	110	110	110	110
P9	12/16	87	87	87	87	87
P10	13/16	104	104	104	104	104
P11	14/16	84	84	84	84	84
P12	15/16	91	91	91	91	91

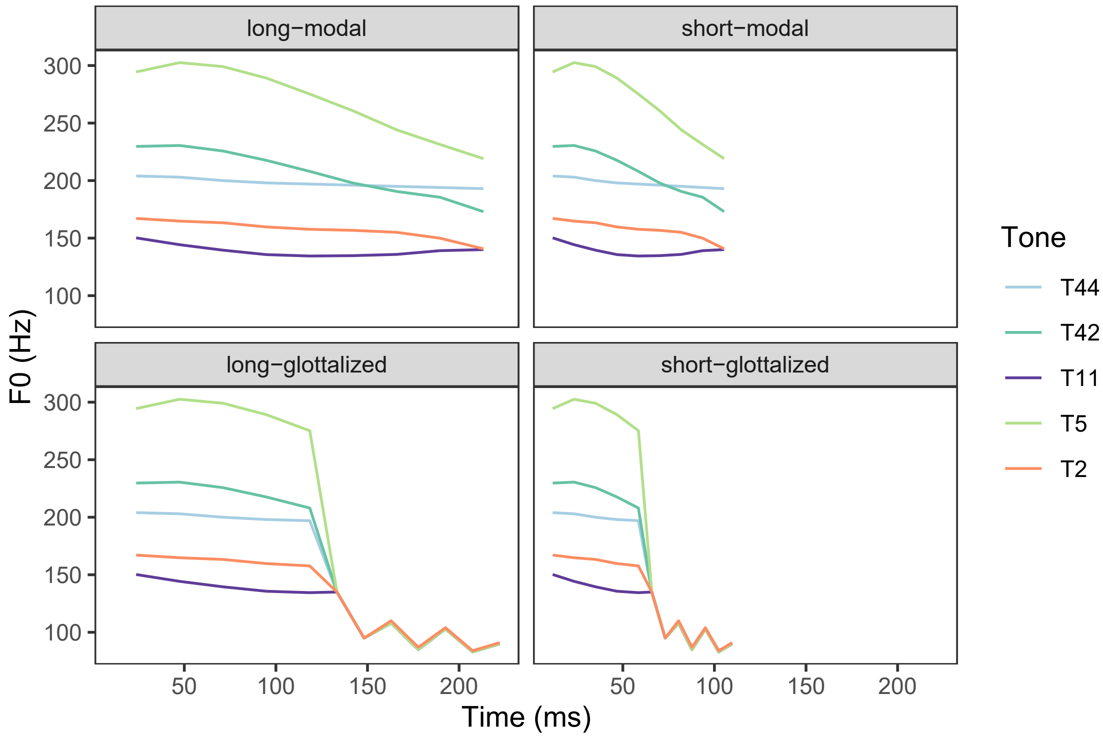


Figure 6.1. The schematized F0 contours for all F0, duration, and voice quality conditions.

The resynthesis was conducted in four steps below. Step 1 and 2 were executed manually. Step 3 and 4 were executed by two Praat scripts (available in Supplementary Material S7 at https://doi.org/10.17605/OSF.IO/3F5MN).

Step 1 Vowel [i] was extracted from the token of [θi 44], and filtered by a Hamming window with a relative width of 1.5. The purpose of using Hamming window is to rescale the intensity of the vowel to a Gaussian shape, which improves the naturalness of the resynthesized tokens.

Step 2 The vowel duration is modified to 240ms and 120ms respectively by deleting pulses in the middle of the vowel where the f0 and formants are steady. The purpose is to create the long and short condition for the duration variable.

Step 3 The F0 of the long and short vowels are modified into the ten different F0 contours in

Tables 6.1 and 6.2 using PSOLA algorithm in Praat using a Praat script.

Step 4 The twenty resynthesized vowels are concatenated with the consonant [θ]. The duration of consonant is modified to 156 ms. The duration of the consonant is the same for both the short and long conditions because the average duration of the onset for checked tone tokens and unchecked tone tokens from the nine speakers in the production experiment is similar to each other (188 ms vs. 176 ms for checked and unchecked tones). [θ] and the resynthesized vowels are concatenated with an overlap of 3ms. The intensity of the fricative is rescaled to 85% of that of the vowel. After the concatenation between the consonant and the vowels, the overall intensity of the concatenated tokens is rescaled to 70dB. A 50ms silence is then added to the beginning and the end of the resynthesized token.

Figure 6.2 present the spectrograms of sample stimuli with High-falling F0 varying in the four conditions of duration and glottalization: High-falling F0 long modal; High-falling F0 short modal; High-falling F0 long glottalized; High-falling F0 short glottalized. I also measured the F0, duration, H1*–H2*, and HNR of all the test stimuli using VoiceSauce (Shue et al. 2011). The results are shown in Figure 6.3. The F0 and duration matches the input value. Tokens in the glottalized condition have lower H1*–H2* and HNR than the modal tokens in the second half of the vowels, suggesting the success of synthesizing creaky constricted and noisy signals, which are very likely to elicit a glottalized percept.

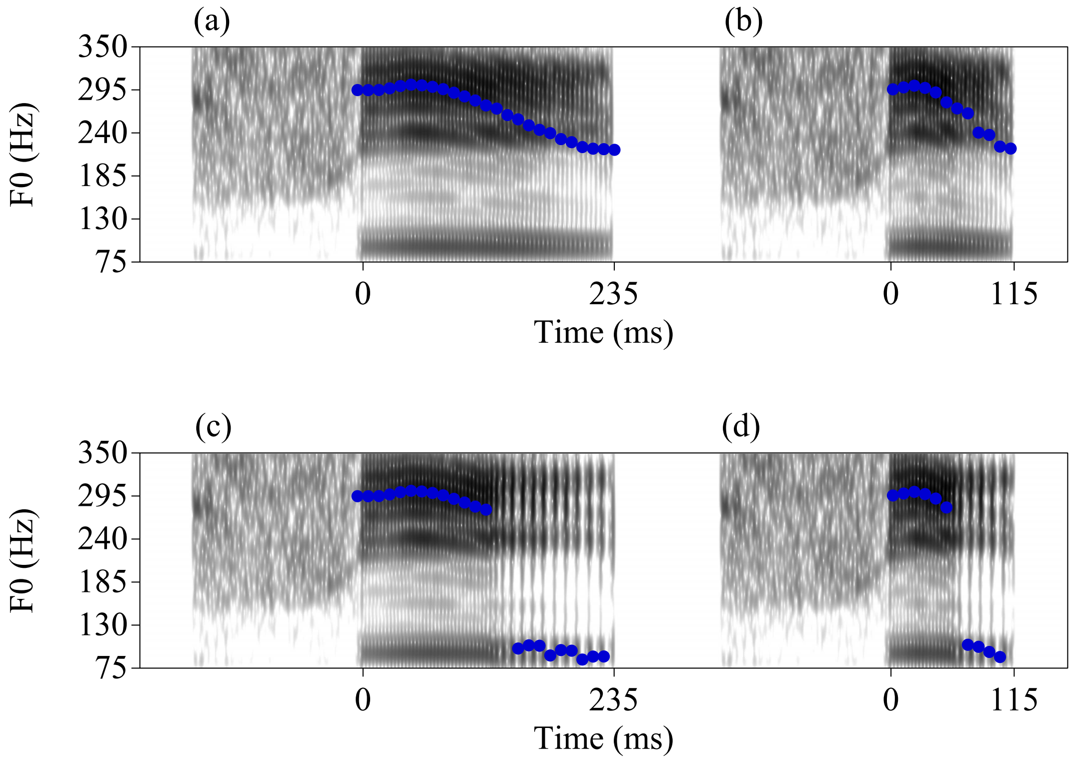


Figure 6.2. Spectrograms of the four conditions of duration and glottalization with High-falling F0: (a): High-falling F0 long modal; (b): High-falling F0 short modal; (c): High-falling F0 long glottalized; (d): High-falling F0 short glottalized.

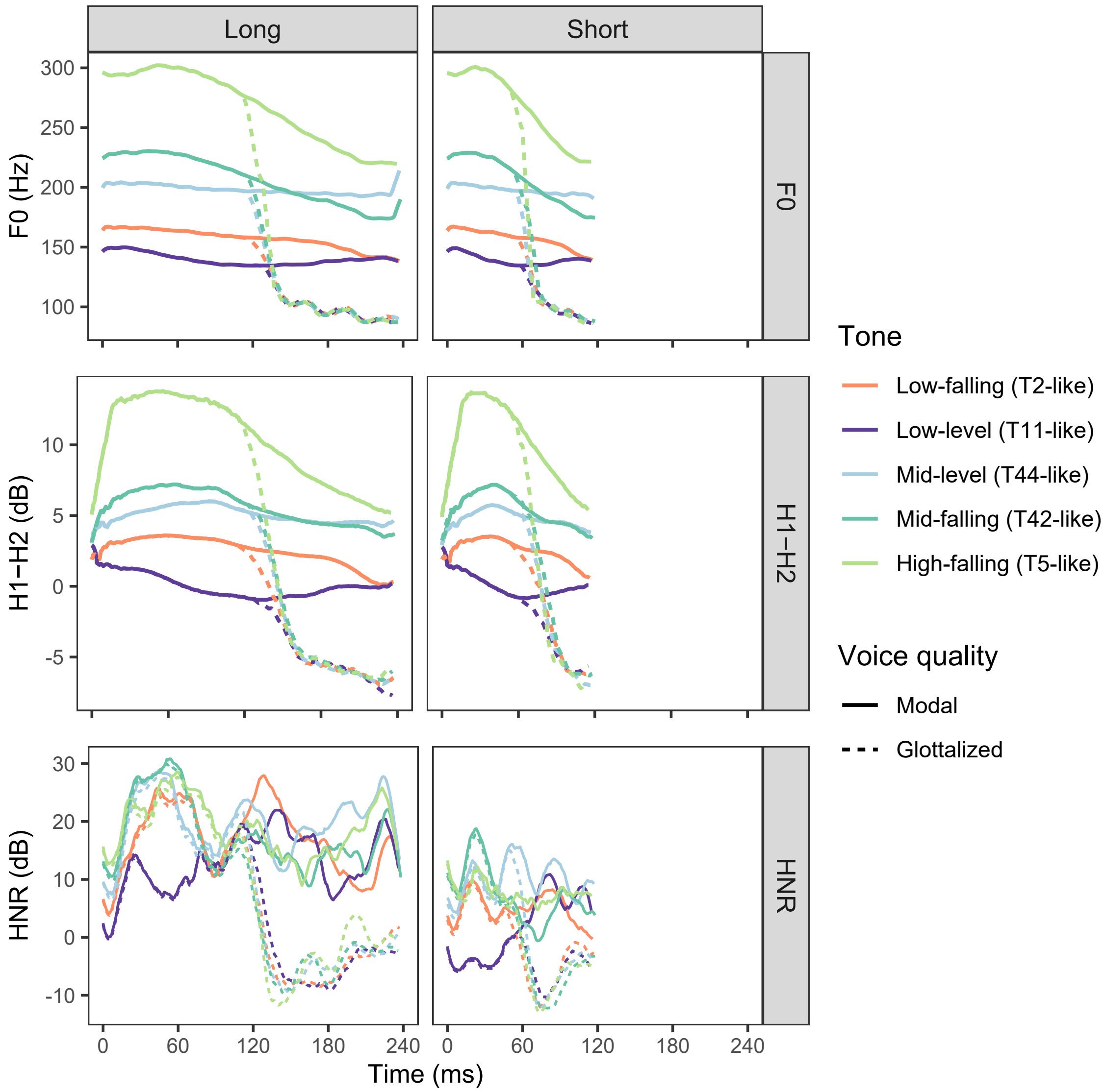


Figure 6.3. F0, H1–H2, and HNR contours of all the test stimuli. For each measurement, there is one data point for every one millisecond. F0 and H1–H2 are in their raw values. The HNR contours are smoothed using "loess" method using the *geom_smooth()* function in the *ggplot2* package in R, because the HNR values have large variation overtime.

There are twenty fillers added to the stimuli. The fillers are resynthesized based on two naturally-produced tokens by Speaker 1: /ti 23/ "地 (ground)" and /ka 35/ "价 (price)". The duration of the base tokens are modified into the same durations as the short and long conditions of the test stimuli. The F0 contour of the two tokens are resynthesized into five rising contours with F0 values between T23 and T35. With the conditions of the three factors fully-crossed with each other (2 segments * 2 durations * 5 F0 contours), twenty filler tokens were created. Rising contours are chosen as stimuli because they are less likely to be confused with checked tones, which have falling F0 contours. The purpose of having those fillers is to distract listeners from knowing that the purpose of the experiment is checked tone identification. Ten additional fillers with rising contours were created based on two naturally-produced tokens by Speaker 1: /te 35/ "替 (substitute)" and /ky 35/ "救 (rescue)". They were used for the practice trial before the start of the experiment to familiarize the listeners with the tasks.

6.2.2 Participant and procedure

Thirty people participated in the experiment (women = 17; men = 13). All thirty participants identified themselves as native speakers of Xiapu Min. Eleven of them also declared Mandarin as their other native language. None of speak languages other than Xiapu Min and Mandarin. Twenty-six of the participants speak only Xiapu Min at home and four of them speak both Xiapu Min and Mandarin at home. One participant had a self-identified hearing disorder, so their results are excluded from all analyses. The average age of the participants is 48 years old (min: 31; max: 68).

The experiment is an identification task and a goodness rating task. The experimenter, who is a native speaker of Xiapu Min, explained the experiment procedures in Xiapu Min to each participant before the start of the experiment, and stayed with the participants during the experiment to answer any questions.

The stimuli consisted of twenty test tokens and twenty filler tokens. There were two blocks in the experiment. The forty stimuli tokens were presented once in each blocks. The order of the tokens within each blocks was randomized for each participant. The experiment was presented through an HTML webpage on a Microsoft Surface Pro 6 laptop with a sound card of Realtek High Definition Audio (SST). The participants listened to the stimuli using a SONY MDR-ZX110AP headset, and gave their responses to the questions using a mouse. Figure 6.4

shows what the participants see during a trial. At the beginning of each trial of the experiment, the stimuli was played automatically to the participants. At the same time, the participants saw two questions on the same screen. The first question asked the participants to choose the word they heard from the seven options. The options are minimal pairs contrasting the seven Xiapu Min lexical tones on /θi/ (See Table 6.3)[1]. The seven options are presented on the screen horizontally aligned with each other. The position of the seven options are randomized by each trial for each participant. The second question asked the participants to rate on a scale of 1–5 in terms of how good the recording fitted the word that they chose[2]. The experiment was self-paced by the participants. The participants could click on the "Replay" button on the webpage to listen to the stimuli as many times as they want. After finishing answering the two questions of each trial, they clicked on the "Submit" button on the webpage to move to the next trial. At the end of the experiment, the participants were asked to produce the seven tonal contrasts in the test trials in a carrier phrase: /wa42 e11 kaŋ42 TARGET tɕja42 ka44 tɕi35/ "我会讲 XX 这个字。(I know how to say the word TARGET.)" Their productions were recorded using the microphone of the same headset for listening. For participants who produced any of the seven target words differently from the forms anticipated in Table 6.3, I discarded all their data from analysis. This is because, if they did not contrast certain tone in their production, their choices in the identification task would not be a correct representation of that tone either. The data from 13 participants were excluded for this reason. Such a high exclusion rate is mainly due to the mispronunciation of the word for T2 "实 (concrete)." This word seldom occurs in monosyllabic forms, but usually occurs in the first position of disyllabic compound words, where tone sandhi applies. Many speaker confused the sandhi form of T2 with its citation form. Data from 17 participants are eligible for the data analysis of the current experiment. The response towards

1. The word for T23 (/θi 23/) was represented by "寺 (temple)" for Participant 1-6 and by "是 (to be)" for the rest of participants because the experimenter and I later found that participants are more familiar with the word "to be" than "temple." Since T23 has a rising F0 contour, which is rather distinct from the checked tones in the F0 space, and T23 is not simulated as one of the F0 conditions in the stimuli, I do not think this change will greatly affect the results of the current experiment.

2. Due to an error in the study design, the first two participants did not have goodness rating results.

the high-falling–short–glottalized token in Block 1 by Participant 14 was excluded from the analysis because the participant informed that they clicked on an unintended option for that trial after the trial had passed. There are 679 data points in total for the result analysis (20 test tokens * 2 repetitions * 17 participants – 1 response error).

Table 6.3. Options of the tonal identification task in the test trials

Tone	Word	Gloss
T2	/θi 2/	"实"
T5	/θi 5/	"湿"
T11	/θi 11/	"时"
T23	/θi 23/	"是"
T35	/θi 35/	"寺"
T42	/θi 42/	"死"
T44	/θi 44/	"诗"

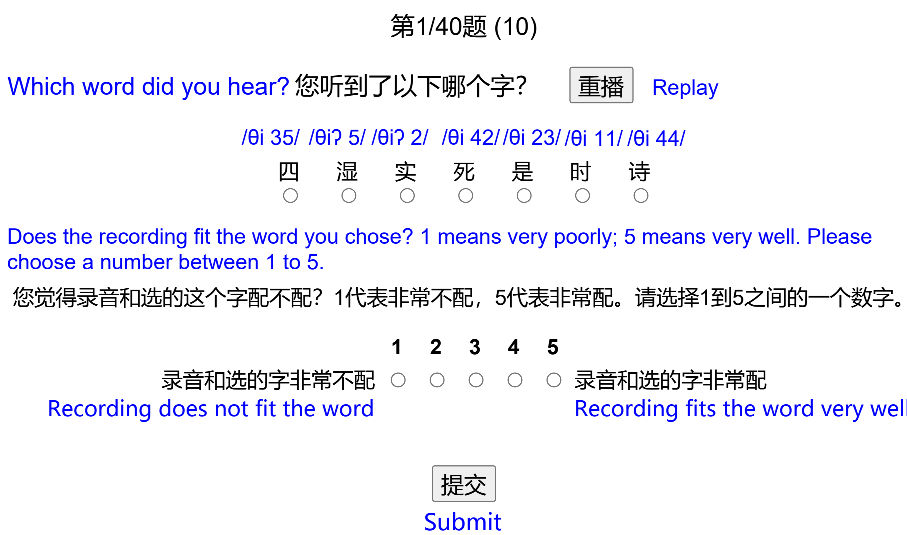


Figure 6.4. A sample page of the test trial for the citation tone identification task. Texts in blue are the translations of the content on the page. The translations were not present during the experiment.

6.3 Results

6.3.1 Descriptive statistics of the identification responses

The R script for the data analysis and the raw data are available in Supplementary Materials S8 and S9 at https://doi.org/10.17605/OSF.IO/3F5MN. The confusion matrix of the participants' responses by the stimuli conditions is in Table 6.4. The percentage of checked T2 and T5 responses in each condition is in Figure 6.5. The average goodness ratings for all stimuli are high (4 and above in 1-5 scale), so they are not very informative and are excluded from the analysis. As Figure 6.5 shows, Checked T2 responses are mostly elicited from tokens with low-falling (T2-like) and low-level (T11-like) F0. The mid-level (T44-like) F0 elicits some T2 responses, but not as many as the former two conditions. Checked T5 are mostly elicited from tokens with high-falling (T5-like), mid-falling (T42), and mid-level (T44) F0. In terms of the duration and voice quality conditions, the short tokens elicit more checked tone responses than the long ones; the glottalized tokens elicited more checked tone responses than the modal ones (except for tokens with mid-level F0 for checked T2 responses). In addition, the effect of glottalization appears to vary by F0 conditions. For the T11-like and T5-like F0 conditions, adding glottalization to a short-modal token does not increase the checked T2 or the checked T5 responses. The ratings are mostly indistinguishable between conditions. All conditions have a mean rating higher than 4 (out of 5).

Table 6.4. Confusion matrix of the responses by condition

Condition \ Response			T2	T5	T11	T23	T35	T42	T44
Low-falling (T2-like)	long	modal	13	0	7	0	0	0	14
		glottalized	17	1	4	0	0	12	0
	short	modal	19	2	7	0	0	1	5
		glottalized	24	2	2	0	0	6	0
High-falling (T5-like)	long	modal	1	15	0	0	0	18	0
		glottalized	1	22	0	1	1	9	0
	short	modal	1	30	0	0	0	3	0
		glottalized	0	29	0	0	1	3	0
Low-level (T11-like)	long	modal	5	0	7	13	0	0	9
		glottalized	19	0	6	2	0	6	1
	short	modal	25	0	4	5	0	0	0
		glottalized	25	1	2	0	1	5	0
Mid-falling (T42-like)	long	modal	0	5	0	1	0	26	2
		glottalized	0	13	0	1	0	20	0
	short	modal	0	17	0	0	0	17	0
		glottalized	2	26	0	0	0	6	0
Mid-level (T44-like)	long	modal	0	0	0	0	0	0	34
		glottalized	3	11	1	1	0	18	0
	short	modal	8	7	1	1	0	0	17
		glottalized	2	21	0	0	0	11	0

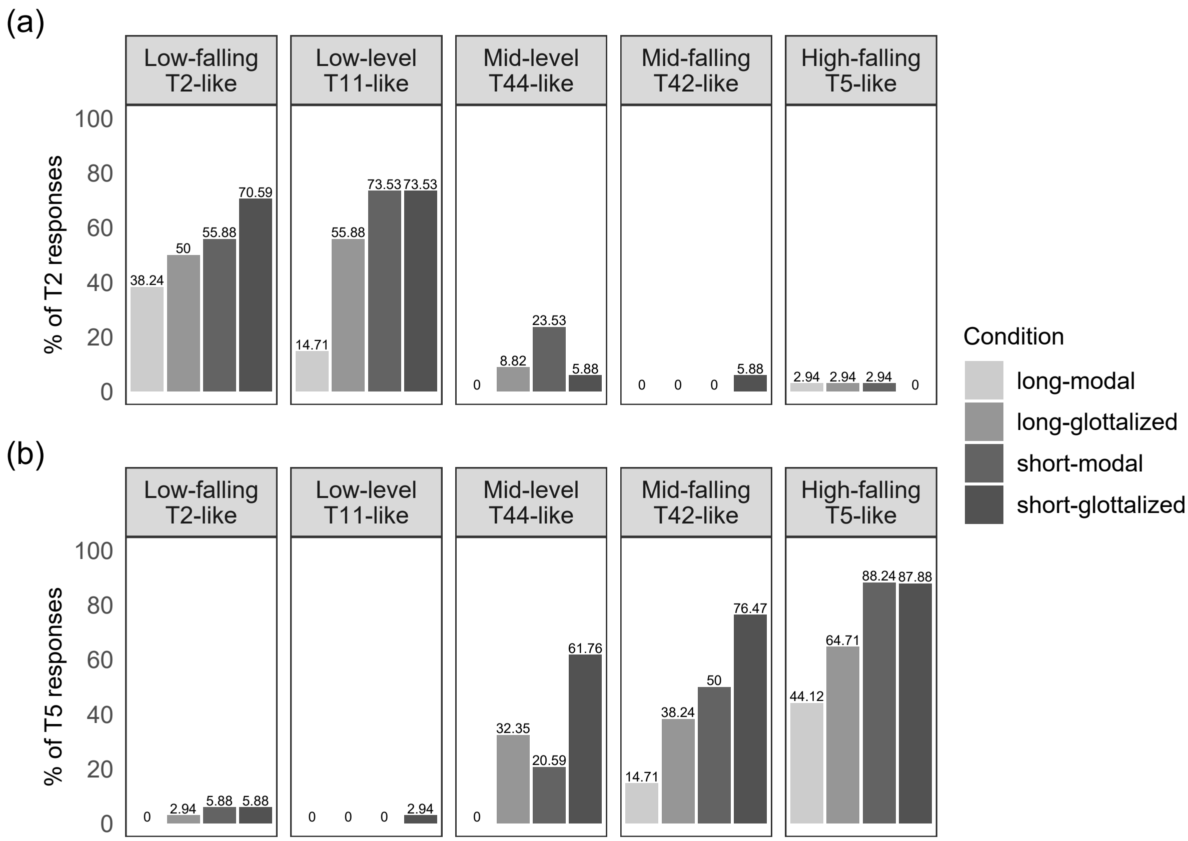


Figure 6.5. Percentage of checked tone responses by conditions. The bars represent the mean percentage. The error bars represent the standard error. A 38.2% T2 response in the low-falling (T2-like) long-modal condition means that, out of all the responses to the stimuli that are long, modal, and have low-falling F0, the percentage of T2 responses is 38.2%. The upper panel (a) shows the percentage of the checked T2 responses by conditions. The bottom panel (b) shows the percentage of the checked T5 responses by conditions. Facets represent the condition of F0. Color represents the conditions of voice quality and duration.

In order to test whether the differences in the likelihood of eliciting checked tone responses between different conditions observed in Figure 6.5 are statistically significant, and whether there are any significant differences between conditions in the goodness rating, I submit the results to linear mixed-effect models. In order to weigh the importance of the three variables – F0, duration, and glottalization – in eliciting checked tone responses, I conduct classification tests using decision trees and Random Forest analysis. In the following sections, I will first present the results of the linear mixed-effect regressions, then the results of the classification tests.

6.3.2 Results from linear mixed-effect regressions

6.3.2.1 Checked T2 responses

The responses from the identification experiment are submitted to linear mixed-effect models using the *glmer* function in *lme4* package (Bates et al. 2015) in R. The R code for the model with checked T2 responses as the dependent variable is in (33), along with the full explanation of the coefficient of each variable. The F0 conditions of low-level, low-falling, mid-level, mid-falling, and high-falling are abbreviated as L, LF, M, MF, and HF in the code. In Model (33), the fixed effects are F0 ($F0_{L\_LF\_M-HF\_MF}$, $F0_{L\_LF-M}$, $F0_{L-LF}$, $F0_{HF-MF}$), $duration$, and $voicequality$ (VQ), the interaction between $duration$ and $voicequality$, and the interaction between $F0_{L-LF}$, $duration$, and $voicequality$. The results of Model (33) are presented in Table 6.5. Across all other conditions, stimuli tokens with a low-level, low-falling, and mid-level F0 elicit more T2 responses than stimuli with high-falling and high-level F0. Within the three former F0 groups, low-level and low-falling F0s elicit more T2 responses than mid-level F0. Low-level F0 does not differ from low-falling F0 in the T2 responses, and high-falling F0 does not differ from mid-falling F0 in the T2 responses.

The interactions between duration and glottalization, and among duration, glottalization, and Low-level vs. Low-falling F0 are significant. The effect of glottalization on eliciting T2 responses is smaller when the vowel duration is short than when the vowel duration is long. Further, the reduction in the effect of glottalization in short tokens is even larger when the token has a low-level F0 than when it has a low-falling F0. The interaction results validates the statistical significance for the phenomenon that, in Figure 6.5, we do not see a difference between short-glottalized and short-modal conditions in the low-level F0 panel, but do see a difference between those two conditions in the low-falling F0 panel.

(33) $$glmer(T2\ response\ or\ not \sim F0_{L\_LF\_M-HF\_MF} + F0_{L\_LF-M} + F0_{L-LF} + F0_{HF-MF}$$
$$+ Duration + VQ$$
$$+ Duration : VQ + F0_{L-LF} : Duration : VQ$$
$$+ (1|Participant),$$
$$family = binomial(link = ``logit''))$$

- $F0_{L\_LF\_M-HF\_MF}$
  - The difference in checked T2 response likelihood between tokens with low-level (L), low-falling (LF), and mid-level (M) F0 vs. tokens with high-falling (HF) and mid-falling (MF) F0
  - $b_1 = \frac{low\text{-}level + low\text{-}falling + mid\text{-}level)}{3} - \frac{(high\text{-}falling + mid\text{-}falling)}{2}$
- $F0_{L\_LF-M}$
  - The difference in checked T2 response likelihood between tokens with Low-level and Low-falling F0 vs. Mid-level F0
  - $b_2 = \frac{low\text{-}level + low\text{-}falling}{2} - mid\text{-}level$
- $F0_{L-LF}$
  - The difference in checked T2 response likelihood between tokens with Low-level vs. Low-falling F0
  - $b_3 = low\text{-}level - low\text{-}falling$
- $F0_{HF-MF}$
  - The difference in checked T2 response likelihood between tokens with High-falling vs. Mid-falling F0
  - $b_4 = high\text{-}falling - mid\text{-}falling$
- $Duration$
  - The difference in checked T2 response likelihood between short tokens vs. long tokens
  - $b_5 = short - long$
- VQ
  - The difference in checked T2 response likelihood between glottalized tokens vs. modal tokens
  - $b_6 = glottalized - modal$

- $Duration : VQ(voice\ quality)$
  - The difference in the effect of glottalization on eliciting T2 between short and long condition
  - $b_7 = (short\ glottalized - short\ modal) - (long\ glottalized - long\ modal)$
- $F0_{L-LF} : Duration : VQ$
  - The difference in the effect difference between adding glottalization to short and long tokens between Low-level and Low-falling F0
  - $$\begin{aligned} b_{low\text{-}level\text{-}duration:VQ} =& (low\text{-}level\ short\ glottalized - low\text{-}level\ short\ modal) \\ & -(low\text{-}level\ long\ glottalized - low\text{-}level\ long\ modal) \\ b_{low\text{-}falling\text{-}duration:VQ} =& (low\text{-}falling\ short\ glottalized - low\text{-}falling\ short\ modal) \\ & -(low\text{-}falling\ long\ glottalized - low\text{-}falling\ long\ modal) \\ b_8 =& b_{low\text{-}level\text{-}duration:VQ} - b_{low\text{-}falling\text{-}duration:VQ} \end{aligned}$$

Table 6.5. Statistics of T2 responses

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-2.792	0.488	-5.724	<.001	***
$F0_{L\_LF\_M-HF\_MF}$	4.426	0.536	8.256	<.001	***
$F0_{L\_LF-M}$	3.675	0.428	8.587	<.001	***
$F0_{L-LF}$	0.038	0.324	0.117	0.907	
$F0_{HF-MF}$	0.436	0.942	0.463	0.643	
$Duration$	1.827	0.312	5.852	<.001	***
VQ	0.880	0.286	3.078	0.002	**
$Duration : VQ$	-1.763	0.572	-3.085	0.002	**
$F0_{L-LF} : Duration : VQ$	-3.141	1.315	-2.389	0.017	*

6.3.2.2 Checked T5 responses

Next, I test the effect of F0, duration, and voice quality on checked T5 responses. The R code for the model with checked T5 responses as the dependent variable is in (34), along with the interpretation of the coefficient of each variable. In Model (34), the fixed effects are F0 ($F0_{HF\_MF\_M-L\_LF}$; $F0_{HF\_MF-M}$; $F0_{HF-MF}$; $F0_{L-LF}$), $duration$, and $voicequality$. The interactions between duration, voice quality, and F0 are non-significant, thus are excluded from in the model. The results of Model (34) are presented in Table 6.6. Across all other conditions, stimuli with a high-falling, mid-falling, and mid-level F0 elicit more T5 responses than those with a low-level and low-falling F0. Within the former three F0 conditions, high-falling and mid-falling F0 elicit more T5 responses than mid-level F0. Further, high-falling F0 elicits more T5 responses than mid-falling F0. Within the latter two F0 groups, low-level F0 does not differ from low-falling F0 in the likelihood of eliciting T5 responses.

(34) $$glmer(T5\ response\ or\ not \sim F0_{HF\_MF\_M-L\_LF} + F0_{HF\_MF-M} + F0_{HF-MF} + F0_{L-LF}$$
$$+ Duration + VQ$$
$$+ (1|Participant),$$
$$family = binomial(link = ``logit''))$$

- $F0_{HF\_MF\_M-L\_LF}$
  - The difference in checked T5 response likelihood between tokens with high-falling, mid-falling, and mid-level F0 vs. tokens with low-level and low-falling F0
  - $b_1 = \frac{high\text{-}falling + mid\text{-}falling + mid\text{-}level)}{3} - \frac{(low\text{-}level + low\text{-}falling)}{2}$
- $F0_{HF\_MF-M}$
  - The difference in checked T5 response likelihood between tokens with high-falling and mid-falling vs. mid-level F0
  - $b_2 = \frac{high\text{-}falling + mid\text{-}falling}{2} - mid\text{-}level$
- $F0_{HF-MF}$
  - The difference in checked T5 response likelihood between tokens with high-falling vs. mid-falling F0
  - $b_3 = high\text{-}falling - mid\text{-}falling$
- $F0_{L-LF}$
  - The difference in checked T5 response likelihood between tokens with low-level vs. low-falling F0
  - $b_4 = low\text{-}level - low\text{-}falling$
- $Duration$
  - The difference in checked T5 response likelihood between short tokens vs. long tokens
  - $b_5 = short - long$

- $VQ(voice\ quality)$
  - The difference in checked T5 response likelihood between glottalized tokens vs. modal tokens
  - $b_6 = glottalized - modal$

Table 6.6. Statistics of T5 responses

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-2.378	0.465	-5.113	<.001	***
$F0_{HF\_MF\_M-L\_LF}$	5.502	0.642	8.566	<.001	***
$F0_{HF\_MF-M}$	2.100	0.320	6.559	<.001	***
$F0_{HF-MF}$	1.887	0.355	5.314	<.001	***
$F0_{L-LF}$	-1.786	1.136	-1.572	0.116	
$Duration$	2.243	0.302	7.432	<.001	***
VQ	1.688	0.285	5.918	<.001	***

6.3.2.3 Summary of results from linear mixed-effect models

To summarize the results from the linear mixed-effect regression tests, I find that short and glottalized tokens have a higher likelihood of eliciting checked tone responses. Whether a low-checked T2 or a high-checked T5 response is elicited depends on the F0 condition. Low F0 (low-level and low-falling) tends to elicit low-checked T2, wheres high F0 (High-falling and Mid-falling) tends to elicit high-checked T5. The mid-level F0 is in between – it can elicit either checked T2 or checked T5, though more T5 responses are elicited on average. In addition, among tokens that are identified as checked T5, tokens with a high-falling and mid-falling F0 on average receive higher goodness rating than those with a mid-level F0. Glottalized tokens receive higher goodness rating than modal ones, indicating that tokens with a high-falling or mid-falling F0, or with glottalization are more prototypical representations of T5. I also find an interaction between duration and glottalization, and between duration, glottalization, and low-level vs. low-Falling F0 in terms of their effect on T2 elicitation: when the duration is short, adding glottalization leads to a smaller increase in T2 response, compared with when the duration is long. When the F0 is low-level, we do not observe an increase from the short-modal condition to the short-glottalized condition (73.53% for both). I propose that, for T2 elicitation, the weakening of the effect of glottalization in short tokens with a low-level F0 is because the cues to glottalization and low F0 overlap with each other and become redundant. Low-falling F0 condition already has a low F0. Glottalization is created by F0-lowering and jittering. Thus, short tokens with low-falling F0 are already a good representation of checked T2. Adding glottalization does not provide new cues to checked T2. Although in the right-most facet of Figure 6.5, we also see that the percentage in checked T5 response does not increase from the short-modal condition to the short-glottalized condition when F0 is high-falling (88.24% vs. 87.88%), the interaction between duration, glottalization, and F0 is non-significant. I propose that adding glottalization does not lead to an increase in the mean percentage of checked T5 response when the token is short and high-falling-toned is due to a ceiling-effect. Tokens with a high-falling F0 and a short duration already elicits checked T5 responses 88.24% of the time. There is not much room for further increases of checked T5 response when glottalization is added.

6.3.3 Classification tests

Knowing that F0, duration, and voice quality each has significant effects on eliciting checked tone responses in Xiapu Min, the next question is, do listeners rely on one cue more than the other? In order to compare the importance of those three factors, we can train classifiers with the factors and the responses from the participants as the input, and have the classifier determine what factors lead to checked tone responses. I choose two algorithms for the response classification: Decision Trees and Random Forests. Decision Trees provide decision rules that decide what tree branches lead to what decision node. The results of Decision Trees can be easily interpreted (Prajwala 2015). Their disadvantage is that it tends to over-fit the data (Prajwala 2015; Ali et al. 2012). Also, during the implementation of the Decision Tree on the current dataset, I found that changing the reference level of the same factor changes the ordering of branching in the Decision rules. This indicates that the branch ordering in the Decision Tree is not a robust representation of the variable importance. Random Forests, in contrast, overcome the over-fitting problem of the decision tree. They also output a more robust and accurate variable importance. In addition, Random Forests produce a higher accuracy for predicting new data, when the training data has a large size (Prajwala 2015; Ali et al. 2012). Their main disadvantage is that it does not output an interpretable decision tree in the same way as the Decision Tree algorithm does. Random Forests simulate a number of trees and evaluate the variable importance based on the overall performance of the trees (Prajwala 2015). Thus, I performed both Decision Tree and Random Forest on the current dataset. The Decision Tree shows what factor(s) lead to checked tone responses. However, the ordering of the factor does not necessarily equal to the importance of the factors. The Random Forest analysis is used to evaluate the importance of the three factors. I used *rpart()* function from *rpart* package for Deicision Tree classification, and used *cforest()* function from *party* package for Random Forest classification. I chose *cforest()* function because this function overcomes the bias towards the factor that have more levels than other factors, which is a problem associated with the widely-used randomForest function (Strobl et al. 2007).

I subset the F0 conditions for checked T2 and checked T5 responses before feeding the data into the classifiers. I only included tokens with F0 of low-level, low-falling, and mid-level when classifying checked T2 responses, and only tokens with F0 of high-falling, mid-falling, and mid-level when classifying checked T5 responses. I excluded the high F0 conditions for T2 responses and low F0 conditions for T5 responses because those conditions elicit near-zero T2 and T5 responses, respectively. Including those unlikely F0 conditions will give a large weight to the F0 factor in the classification, making the classification results less informative. Thus, I only included tokens with an F0 that is potentially confusable with the F0 of checked T2 and checked T5, then compare the importance between F0, duration, and voice quality in terms of eliciting checked tone responses. I will present the classification results for checked T2 responses first, checked T5 responses next. Within each response, I present the result for the Decision Tree first, the results for Random Forest next.

The procedure for Decision Tree classification is that first, I partition the data into a training set and a test set with a proportion of 80% and 20% of the dataset. The Decision Tree is first trained on the training set using ten-fold cross-validation. The cross-validation outputs a Complexity Parameter (cp) that yields the highest prediction accuracy in the cross-validation. The parameter – cp – defines the minimum required improvement for the tree to further grow. If cp is 0.01, and a branching improves the accuracy of the classification less than 0.01, then the branching is stopped. After a cp value is outputted from the cross-validation, I run the Decision Tree again using that cp value on the whole training data. The output decision tree is the final decision tree. The final model calculated on the whole training set is used to estimate the test set. The accuracy of the classification of the test set is calculated. The R code for the Decision Tree model is in (35).

(35) $$\begin{aligned} rpart(checked\ or\ not \sim F0 + Duration + VQ,&\\ data = training,&\\ method = \text{``}class\text{''},&\\ control = rpart.control(cp = XX))& \end{aligned}$$

Random Forest classification uses the same training and test sets as the Decision Tree classification. The Random Forest model is built using the training set. The R code for the Random Forest model is in (36). The hyper-parameters for the *cforest()* function is set as suggested by Strobl et al. 2007, to avoid the bias towards factors that have multi-level categories. The Random Forest classification outputs the variable importance for each factor. A larger number represents that the variable weighs more in the classification. The classification model built upon the training set is then fit on the test set. The accuracy of the classification on the test set is calculated.

(36) $cforest(checked\ or\ not \sim F0 + Duration + VQ,$

$data = training,$

$controls = cforest\_control(teststat = ``quad",$

$testtype = ``Univ", mincriterion = 0,$

$ntree = 50, mtry = 3,$

$replace = FALSE))$

(Parameters in *cforest_control* are from the supplementary materials of Strobl et al. 2007.)

6.3.3.1 Classification results for Checked T2 responses

First, in terms of the classification of checked T2 responses, the decision tree is in Figure 6.6. The cp value is 0.0234375, as outputted from the cross-validation tests. The prediction accuracy using the final model on the test set is 71.6% (chance level: 60.49%; p = .025). According to the decision tree, the three F0 conditions are divided into two categories: mid-level vs. low-level and low-falling. Tokens with a low-level or low-falling F0, short duration; or with low-level or low-falling F0, long duration, and glottalization lead to checked T2 responses. Random Forest ranks the importance of the three variables as F0 (0.142) > Duration (0.062) > Voice quality (0.025). The prediction accuracy using the Random Forest model on the test set is 71.6% (p = .025), same as the prediction accuracy using the Decision Tree model.

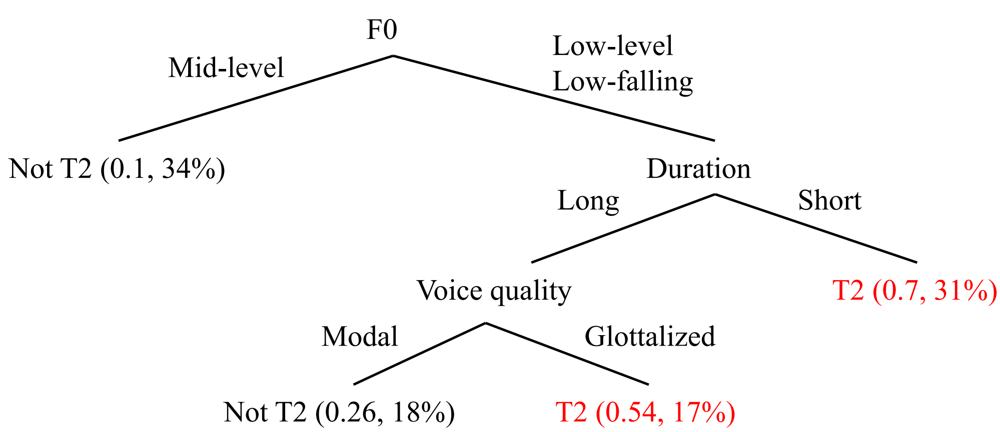


Figure 6.6. Decision tree for what parameters lead to checked T2 responses. The two numbers in the parenthesis after each end node are the probability of that branch is identified as T2, and the percentage of that branch in the data.

6.3.3.2 Classification results for Checked T5 responses

Second, in terms of the classification of checked T5 responses, the decision tree is in Figure 6.7. The cp value is 0.00955414, as outputted from the cross-validation tests. The prediction accuracy using the final model on the test set is 65.43% (chance level: 51.84%; p = .009). According to the decision tree, the three F0 conditions are divided into two categories: high-falling vs. mid-falling and mid-level. Tokens with high-falling F0; or with mid-falling or mid-level F0, short duration, and glottalization lead to checked T5 responses. Random Forest ranks the importance of the three variables as Duration (0.102) > F0 (0.091) > Voice quality (0.066). However, the importance for Duration and F0 is similar. I ran the Random Forest model with twenty different random seeds. Eleven out of twenty times, duration ranks higher than F0; whereas in the other nine times, F0 outranks duration. Voice quality ranks lowest in all twenty tests. Based on the Random Forest results, Duration and F0 have near-equal importance in T5 response classification. Voice quality has lower importance than duration and F0. The prediction accuracy using the Random Forest model on the test set is 74.85% ($p < .001$), which is higher than the prediction accuracy using the Decision Tree model.

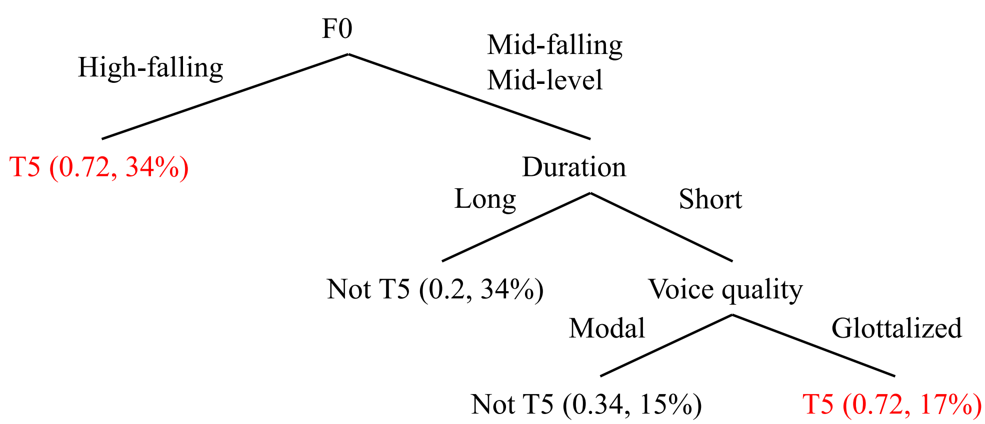


Figure 6.7. Decision tree for what parameters lead to checked T5 responses. The two numbers in the parenthesis after each end node are the probability of that branch is identified as T5, and the percentage of that branch in the data.

I summarize the conditions that lead to a checked tone responses in the Decision Tree in

Table 6.7. For checked T2 responses, when F0 is mid-level, the response is always classified as non-T2, regardless of the duration or glottalization conditions. When stimuli have a low-level or low-falling F0, and are short, the responses are classified as T2, regardless of the voice quality condition. For checked T5 responses, when F0 is high-falling, the responses are classified as T5, regardless of the duration or voice quality conditions. When F0 is mid-falling or mid-level, only the short and glottalized tokens are classified as T5. The Decision Tree results confirms our observation from the linear mixed-effect models: Low-level and low-falling F0 elicit more low-checked T2 responses than other F0 conditions; and high-falling F0 elicits more high-checked T5 responses than other F0 conditions. Short and glottalized tokens are more likely to elicit checked responses (both low-checked T2 and high-checked T5). Regarding the importance of the three variables – F0, Duration, and Voice quality – in terms determining whether the response will be checked or not, F0 (low-level, low-falling, and mid-level) is the most important variable, followed by duration and voice quality for checked T2 classification. For checked T5 classification, F0 (high-falling, mid-falling, and mid-level) and duration have similar importance, but are both more important than voice quality. For both checked T2 and T5 classification, voice quality has the lowest importance. The results from Random Forest also confirm the mixed-effect models and the distribution of checked tone responses under each condition. We see that adding glottalization to short tokens does not increase the percentage of checked T2 responses under the low-level F0 condition, or the percentage of checked T5 responses under the high-falling F0 condition. The results from Random Forest verifies the relatively lower importance of voice quality in eliciting the checked tone responses.

Note that the importance scores from the Random Forest models are relative, and only apply to the specific conditions used in the current experiment. A rank of F0 > Duration > Glottalization does not imply that F0 is always more important than the other two factors in eliciting checked tone response. The F0 conditions used in the current conditions differ in both F0 heights and contours. The variation between conditions is not evenly-spaced. If the F0 conditions are closer-spaced to each other, and are very similar to the F0 values of the checked tones, it is possible that the importance of F0 decreases. The duration variable only have two levels – 120 ms and 240 ms. If there are multiple levels of duration, the importance of duration might change as well. Lastly, glottalization also only have two levels – modal or having glottalization at the latter half of the vowel. If the proportion and the timing of the glottalization in the vowels vary in more levels, the classification results might change. Moreover, the glottalization in the current stimuli is created by lowering and jittering the F0. Such kind of glottalization resembles the prototypical type of creak (Keating, Garellek, and Kreiman 2015). There are other types of creaky voice – vocal fry, tense, irregular, pitch doubling – which are not tested in the current study. Study has shown that listeners are sensitive to different types of creaky voice in tone identification (Huang 2020). It is possible that the listeners are not sensitive to the prototypical type of creak in the stimuli as a cue to checked tone, but will use other types of creaky voice as a cue.

Table 6.7. Branches leading to T2 and T5 in the Decision Trees

	F0	Duration	Voice quality
Checked T2	Low-level, Low-falling	Short	
	Low-level, Low-falling	Long	Glottalized
Checked T5	High-falling		
	Mid-falling, Mid-level	Short	Glottalized

6.4 Summary

In order to test whether listeners' of Xiapu Min use cues of F0, duration, and voice quality, which are three cues found to differentiate checked tones from unchecked tones in the production of Xiapu Min, a tone identification task has been conducted. Stimuli were resynthesized into five F0 conditions, long and short durations, and modal and glottalized voice quality. The results show that listeners make use of all three cues when identifying checked tones. Tokens with a low F0 (low-falling and low-level) elicit more low-checked T2 responses, whereas tokens with a high F0 (high-falling) elicits more high-checked T5 responses. Short and glottalized tokens elicit more checked tone responses for both T2 and T5. Among the three factors, glottalization is of the lowest importance in eliciting checked tone responses. Future studies should create more levels for each parameter. Adding more evenly-spaced levels to the F0 and duration parameters can help illustrate whether the effect of F0 and duration is linear or categorical. The glottalization factor can be modified by varying the degree, the length, the timing, and the type of glottalization. It will enable us to test whether Xiapu Min listeners are sensitive to more fine-grained cues of glottalization when identifying checked tones.

Chapter 6, in full, is currently being prepared for submission for publication of the material. Chai, Yuan, and Shihong Ye. "Acoustic cues for checked tone perception in Xiapu Min". The dissertation author was the primary investigator and author of this material.

Chapter 7

Perception of Xiapu Min checked tones in sandhi forms

7.1 Introduction

In Chapter 6, I tested which cues that listeners use when identifying Xiapu Min checked tones in citation forms. In the current chapter, I focus on the perception of checked tones in sandhi forms. As discussed in Chapter 5, Xiapu Min checked tones undergo tone sandhi processes when they are in the first position of compound words, and are neutralized with unchecked tones phonologically. Checked T2 is neutralized with T23 and T44 into a mid-level T44. Checked T5 is neutralized with T35 and T42 into a high-level T55 (Example (37); see Table 5.1 for examples). The results from Chapter 5 show that in sandhi forms, checked T2 and checked T5 are primarily distinguished by shorter duration, though the duration differences between underlying checked and unchecked tones after sandhi are substantially reduced. In the current chapter, I posit the question: are Xiapu Min listeners able to differentiate checked tones from unchecked tones in their neutralized sandhi forms? If so, what cue(s) do they use to differentiate the neutralized forms? Since in production, duration is the primary correlate of checked tones in sandhi forms, I hypothesize that if listeners are able to discriminate checked tones vs. unchecked tones in sandhi forms, the main perceptual cue that they use will also be duration.

(37) a. /T2, T23, T44/ → [T44] / ___ X
b. /T5, T35, T42/ → [T55] / ___ X

Studies have found that phonological neutralization can be either complete or incomplete at the phonetic level, and this need to be accounted for from two perspectives: production and perception. Dinnsen (1985) proposed four types of neutralization in terms of their phonetic completeness in production and perception. In Table 7.1, I present the four types proposed by Dinnsen (1985) and summarize the examples of each neutralization type.

Table 7.1. Four types of neutralization in production and perception (Dinnsen 1985, p. 274, Table 2) and examples

Type	Production differences	Perceptual differences	Neutralization between
A	No	No	• Korean /t/–/t^h/–/s/ • Taiwanese Min T2–T5 • Fuzhou Min T44–T242–T53
B	Yes	No	• Russian voiced–voiceless stops • Polish voiced–voiceless stops • English intervocalic /t/–/d/ • Mandarin T3–T2
C	Yes	Yes	• German voiced–voiceless stops • Eastern Andalusian Spanish /s/–/p/–/k/
D	No	Yes	• Dutch voiced–voiceless stops • Contextual influences on English vowel perception • Contextual influences on Mandarin tone perception

Neutralization of Type A is phonetically complete in both production and perception. No phonetic difference is found in the production of the neutralized contrasts, and listeners cannot differentiate the contrasts in neutralized forms. Dinnsen (1985) maintained that Type A neutralization is not well-established because they did not find a case where the neutralization is complete in production among the eight empirical cases that they examined in the study. They also highlighted that Type A neutralization is problematic because it is nearly empirically impossible to measure every phonetic aspect of production and every acoustic cue for perception to ensure that no differences exist in both dimensions. Despite the potential empirical difficulties, several studies have reported complete neutralization in production and perception in terms of the specific phonetic dimensions that they chose to measure. Kim and Jongman (1996) found that the Korean /t/–/tʰ/–/s/ coda neutralization to [t] is phonetically complete in production (See Examples in (38)). There is no difference in the duration of the [t] closure, or of the preceding vowel between words with underlying /t/, /tʰ/, and /s/ codas. The results from an identification task showed that the identification accuracy of words with the neutralized forms of /t/, /tʰ/, and /s/ is at chance (32%), suggesting a complete neutralization in perception. Another example of complete production and perception neutralization is the tone neutralization resulted from tone sandhi in Taiwanese Min. Chien and Jongman (2019) found that the neutralization between checked T2 and T5, unchecked T51 and T55, and T33 and T21 is complete in production. There is no difference in F0 height, F0 contour shape, or duration. Kuo (2013) found that in the identification task of citation and sandhi T2 (realized as [T5]), and citation and sandhi T5 (realized as [T2]), the listeners are more likely to choose the surface form of the sandhi tones (i.e. choose [T5] for sandhi /T2/ and [T2] for sandhi /T5/). Moreover, the percentage of choosing [T5] for sandhi /T2/ is even higher than choosing [T5] for citation /T5/. The results from Kuo (2013) indicate that Taiwanese Min listeners are not able to perceptually differentiate tokens with the same surface tone but different underlying tones. The last example of complete perception and production neutralization is from Fuzhou Min. In Fuzhou Min, T44, 242, and 53 are neutralized into T44 before T53 (Li 2015). No difference has been found in the pitch height among those tones in their sandhi forms, and the classification accuracy of those three tones is not above chance in identification task.

(38)	/kət/	"to collect"	/kət + ʧi/	[kət.ʧ'i][a]	"to collect" + NEG
	/kətʰ/	"outside"	/kətʰ + kwa/	[kət.k'wa]	"outside" + "and"
	/kəs/	"thing"	/kəs + kwa/	[kət.k'wa]	"thing" + "and"

(Kim and Jongman 1996, p. 298)

[a] Kim and Jongman (1996) called consonants with ' as "reinforced consonants" in Korean.

Neutralization phenomena of Type B are phonetically incomplete in production, but complete in perception. Such kinds of neutralization indicate that a sound change is in progress –

while there remain differences in the production between two contrasting sounds, listeners are not longer sensitive to those differences in the perception. The contrasts are in the process of neutralization. Dinnsen (1985) maintained that Type B is also empirically problematic because when an experiment shows that the listeners do not differentiate the differences in production, it is possible for the experiment design to not be able to capture the perceptual sensitivity. Other studies have provided examples of Type B. Matsui (2011) found that in Russian, vowels preceding an underlyingly voiced coda have a longer duration than those preceding an underlyingly voiceless coda, indicating that neutralization of phrase-final stop devoicing is incomplete in production. In perception, the accuracy of identifying words with voiced and voiceless coda is marginally above chance (54%), and there is a strong bias towards the voiceless option. This suggests that Russian listeners are not sensitive to the vowel length cue when determining the underlying voicing of the coda. Similar results have been found for the neutralization between voiced and voiceless stop in Polish (Slowiaczek and Szymanska 1989). Another example is the neutralization of English /t/ and /d/ into flap [ɾ] in intervocalic position. Braver (2014) reported that vowels preceding /d/ flaps are lightly longer than before /t/, which indicates an incomplete neutralization of /t/ and /d/ in intervocalic position. However, in a 2AFC (two-alternative forced-choice) perception task, the listeners were not able to identify /d/ and /t/ accurately, suggesting that the perceptual neutralization between /t/ and /d/ flaps is complete. Another example is in the suprasegmental level: Mandarin T3-T2 neutralization. In Mandarin, T3 becomes T2 before another T3 (T3 $\rightarrow$ T2 / ____ T3), resulting in T3-T2 neutralization before T3. The neutralization between T3 and T2 is incomplete in production in that sandhi T3 has lower F0 height and less F0 rising than T2 (Yuan and Chen 2014; Peng 1996, 2000; Kuang 2018). The perceptual neutralization between sandhi T3 and T2 is complete in that the speakers were unable to distinguish sandhi T3 from citation T2 either in isolation (Peng 1996) or in compound (i.e. T3-T3 vs. T2-T3 sequence) (Peng 2000).

Type C neutralization has incomplete neutralization in both production and perception. Dinnsen (1985) provided an example of Type C – German final obstruent devoicing. Voiced obstruents are devoiced and neutralized with voiceless obstruents in word-final position. Despite the neutralization, underlyingly voiced obstruents still differ from underlyingly voiceless obstruents in terms of preceding vowel length, percentage of voicing during the closure, and aspiration duration. Listeners are able to differentiate voiced obstruents from voiceless obstruents in neutralized forms above chance. Roettger et al. (2014) replicated the study and confirmed that in German, the vowels before devoiced stops are significantly longer than those before voiceless stops. The listeners were able to differentiate devoiced stops from voiceless stops, but the accuracy of discrimination is barely above-chance (55%). Besides the German final stop devoicing example, Eastern Andalusian Spanish and Mizo also have Type C neutralization phenomena. In Eastern Andalusian Spanish, obstruent codas are realized as aspiration. When the obstruent codas are in word-medial position, the following consonant is also lengthened (Example (39)). Bishop (2007) found that, on one hand, the aspiration from /s/ is longer than the aspiration from /p/ and /k/. On the other hand, the stop following /p/ and /k/ has a longer period of closure than the stop following /s/. They conducted an identification task where the duration of aspiration and stop closure varied along a continuum. Listeners were more likely to perceive /p/ when the closure of the following stop was longer. The changes in the aspiration duration did not have a significant effect on their responses. This indicates that listeners of Eastern Andalusian Spanish are sensitive to the durational difference in stop closure following /p/ and /s/, but not the durational difference of the aspiration that replaces /p/, /k/, and /s/. The last example is from Mizo, where rising tone becomes low tone when followed by a high or falling tone, resulting in a neutralization between the sandhied rising tone and the citation low tone (Lalhminghlui and Sarmah 2018). The production results showed that the citation low tone has a falling F0 contour, while the low tone derived from rising tone has a level F0 contour. The average F0 of the rising tone-derived low tone is also higher than the citation low tone. The identification task results show that when presented with a surface low tone, the listeners are able to identify whether its underlying tone is a rising tone (65%) or a low tone (60%) significantly above chance.

(39) /as.ta/ [a^h.tːa] "until"
/ap.ta/ [a^h.tːa] "apt"
/ak.ta/ [a^h.tːa] "certificate"
(Bishop 2007, p. 1765)

For neutralization phenomena of Type D, there are no phonetic differences in production, but the listeners are still able to differentiate the neutralized contrasts in perception. Dinnsen (1985) claimed that this type of neutralizaiton is logically impossible because if there is no difference in the source of perception, it is impossible to initiate perceptual difference. However, Jongman (2004) and Warner et al. (2004) have shown that listeners are capable of using cues created in the resynthesized signals that are not present in the natural speech to discriminate neutralized segments. In Dutch, voiced and voiceless stops are neutralized into voiceless stops in phrase-final positions, but distinct in word-medial positions (40). In production, there is no significant difference in the closure duration between voiced and voiceless stops in the neutralization environment. But in word-medial positions, voiceless stops have longer stop closure than voiced stops. Jongman (2004) and Warner et al. (2004) tested whether changes in stop closure duration would affect listeners' identification of the voiced and voiceless stops in phrase-final position, despite the lack of differences in stop closure duration in that environment in the production. They varied the duration of the stop closure of /t/ and /d/ on a 11-step continuum. The longest duration was 100 ms longer than the shortest duration. The results showed that listeners gave more /t/ responses for longer closure duration when identifying voiced vs. voiceless stops in the neutralization environment. The studies by Jongman (2004) and Warner et al. (2004) indicate that listeners are able to use cues (i.e. closure duration) that are not found in natural production to distinguish neutralized forms in resynthesized signals. In the case of Dutch voiced vs. voiceless stop perception, the listeners resort to the closure duration differences found in unneutralized word-medial positions, when they are distinguishing those stops in neutralized phrase-final positions.

(40) Dutch: Phrase-final position; voicing neutralization

/bɔt/	[bɔt]	*bot* "bone"	/bɔd/	[bɔt]	*bod* "offer"

Word-medial position; voicing preserved

/wɛtən/	[wɛtən]	*wetten* "laws"	/wɛdən/	[wɛdən]	*wedden* "to bet"

(Warner et al. 2004, p. 256)

Another example of no acoustic difference being present but the perception being different is that the perception of stimuli can be influenced by the contextual information. Even if the stimuli is the same, the listeners may give different responses when the context is different. Ladefoged and Broadbent (1957) found that when the F1 of the vowels in preceding sentence is lowered, a word that is originally perceived as "bit" is perceived as "bet" after the F1 lowering. When the F1 of the vowels in preceding sentence is raised, a word that is originally perceived as "bet" is then perceived as "bit" after the F1 raising. Similar contextual normalization phenomenon has been observed in tone identification as well. Huang and Holt (2009) found that for Mandarin listeners, when a target word was preceded by a sentence with a higher mean F0, more T2 (35) responses were elicited; whereas when a target word was preceded by a sentence with a lower mean F0, more T1 (55) responses were elicited. This was true for both speech and non-speech contexts. Thus, although Dinnsen (1985) claimed Type D neutralization to be impossible, empirical studies have proved that people do perceive differences despite they are not present in the production. This can be due to listener's accessing their knowledge of the unneutralized underlying forms, or the influence of contexts.

Chapter 5 has shown that in the sandhi forms of T2, 44, and 23, checked T2 can be significantly distinguished from T44 and T23 by duration; in the sandhi forms of T5, 42, and 35, checked T5 can be significantly distinguished from T42 and 35 by duration and HNR. Thus, the neutralization between checked tones and unchecked tones will belong to either Type B or C of the neutralization typology (Dinnsen 1985). If listeners are able to differentiate checked tones from unchecked tones in perception, the neutralization of checked and unchecked tones will belong to Type C; if not, their neutralization will belong to Type B. In order to find out which type of neutralization the checked–unchecked tone neutralization in Xiapu Min belongs to, I conducted a identification task, in which the listeners listened to the natural stimuli that have the same surface tone in sandhi forms but differ in citation forms, then identify the word that they hear. The purpose is to see whether listeners can successfully identify the words above chance, and whether their answers are different when the underlying tone is different.

7.2 Stimuli

The stimuli used in the current experiment are available in Supplementary Material S10 at https://doi.org/10.17605/OSF.IO/3F5MN. The neutralizing pairs that are tested here include T2–T44，T23–T44, T5–T35, T5–T42, and T35–T42. The contrast between T2–T23 was not tested because I did not find minimal pairs that suited the stimuli selection criteria. The target words are presented in Table 7.2. Within each pair, the target syllables are the first syllable of each compound. The target syllables have the same segmental structure and the same surface tone, but differ in their underlying tone. The onset of the second syllable of the compounds has the same tone, place of articulation, and sonorancy, in order to minimize any potential confounding effects of the second syllable on the target syllable. One exception is the stimuli for the contrast T35–T42, in which the onsets of the second syllable following T35 and T42 differs in their place of articulation (/k/ vs. /θ/). However, I measured the formants of T35 and T42 in both compounds, and their F2 values are rather similar.

Table 7.2. Stimuli of sandhi tone identification task

Contrast	Segment		Compound		
T23–T44	/to 23/	路	/to 23 keu 42/	[to 44 keu 42]	路口 "road corner"
	/to 44/	刀	/to 44 keu 42/	[to 44 keu 42]	刀口 "knife scar"
T2–T44	/tsa 2/	杂	/tsa 2 ki 44/	[tsa 44 ki 44]	杂技 "acrobatics"
	/tsa 44/	查	/tsa 44 kaŋ 44/	[tsa 44 kaŋ 44]	查岗 "change for absence"
T5–T35	/tʰe 5/	铁	/tʰe 5 pain 42/	[tʰe 55 pain 42]	铁板 "iron plate"
	/tʰe 35/	替	/tʰe 35 po 42/	[tʰe 55 po 42]	替补 "substitute"
T5–T42	/θi 5/	湿	/θi 5 ti 35/	[θi 55 ti 35]	湿地 "wetlands"
	/θi 42/	死	/θi 42 tsui 35/	[θi 55 tsui 35]	死罪 "death penalty"
T35–T42	/ka 35/	价	/ka 35 kai 42/	[ka 55 kai 42]	价格 "price"
	/ka 42/	假	/ka 42 θe 42/	[ka 55 θe 42]	假设 "hypothesize"

Because the compounds used in the stimuli generally differ in terms of the second syllable (except in the T23–T44 contrast), the second syllable is replaced by a period of noise. The compounds are cut off at the end of the first syllable, which is when the formant amplitude drops or when the voicing stops, whichever comes first. A 342-ms interval of pink (speech-shaped) noise of 342 ms generated in Audacity (Audacity team 2022) and concatenated to the first syllable's cut-off point. The amplitude of the noise is scaled to 60% of that of the first segment. After the concatenation, the amplitude of the modified compound is rescaled to 70 dB. Last, 50 ms of silence was padded to the beginning and the end of the compound. The original recordings come from the natural productions of two female Xiapu Min speakers (Speaker #1 and Speaker #9 in the production study in Chapters 4 and 5). They produced two tokens for each target word in Table 7.2. One exception is that, for the word /tʰe 35 po 42/ 替补 "substitute" in the T5–T35 contrast, Speaker #9 only produced one token. In order to ensure a balance of the data, I duplicated that word from Speaker #9 in the stimuli. In total, there are 40 tokens in the stimuli for the compound identification task (5 contrasts * 2 words in contrast * 2 speakers * 2 tokens). Sample spectrograms for the T5–T35 contrast are in Figure 7.1.

The F0, duration, H1*–H2*, and HNR of the stimuli were calculated in VoiceSauce (Shue et al. 2011) and are presented in Figure 7.2 and Table 7.3. F0 is transformed into semitones, with the mean value of F0 of all tokens in the stimuli as the baseline, because semi-tone is closer to the listener's perception of pitch. The descriptive statistics of the stimuli show that the F0 and H1*–H2* of the tones within each neutralization pair are similar. In terms of duration, T2 and T23 have a shorter duration than T44. T5 has a shorter duration than T35 and T42. In terms of HNR, T5 has a lower HNR than T35. The acoustic differences in neutralized sandhi forms reported in the production study in Section 5.2 are also present in the stimuli selected for the current experiment.

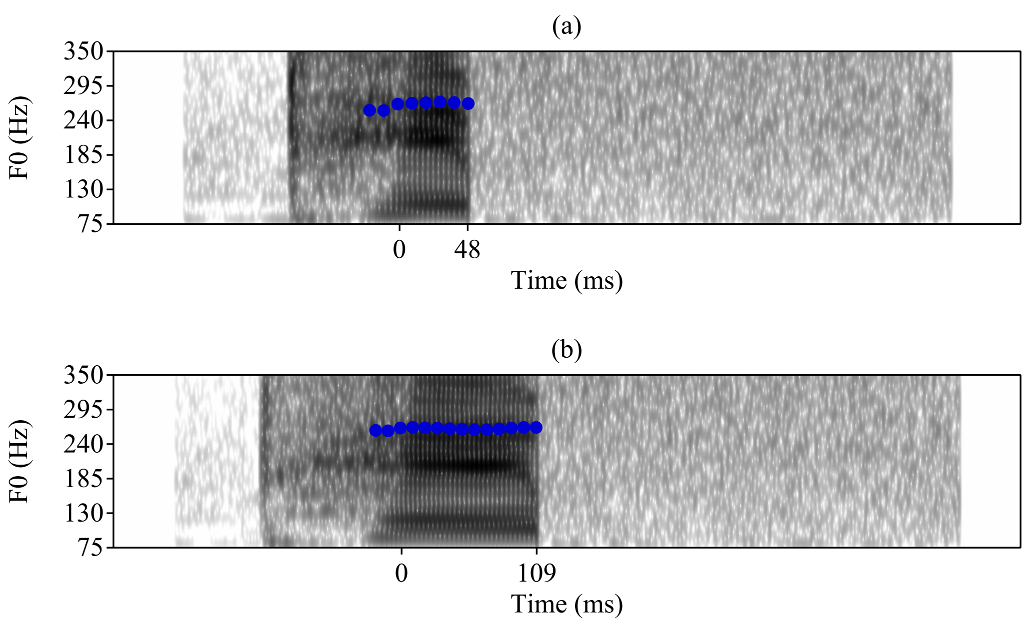


Figure 7.1. Sample spectrograms of the stimuli for T5–T35 contrast. (a) /tʰe 5 pain 42/ 铁板; (b) /tʰe 35 po 42/ 替补

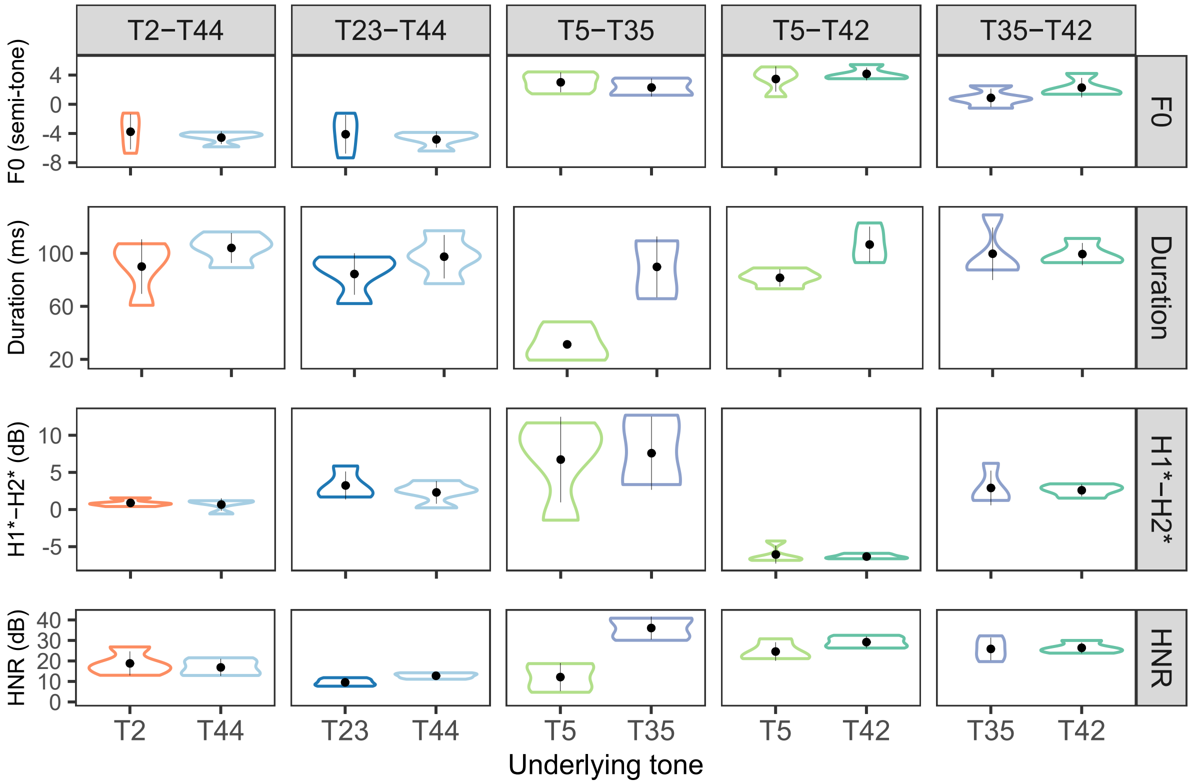


Figure 7.2. Stimuli of the neutralization experiment. The upper and lower whiskers extends to the largest and smallest values in the data, but not further than the 1.5 times of the box length.

Table 7.3. Mean F0 (Hz), F0 (semi-tone), H1*–H2*, HNR, and duration of the stimuli of the neutralization experiment

Contrast	Underlying tone	F0 (Hz)	F0 (semitone)	H1–H2	HNR	Duration
T2–T44	T2	199.594	−3.207	0.891	18.802	90.007
	T44	189.324	−4.569	0.668	16.849	104.05
T23–T44	T23	195.862	−4.109	3.238	9.555	84.39
	T44	186.645	−4.821	2.304	12.762	97.458
T5–T35	T5	293.499	3.008	6.727	12.133	31.262[c]
	T35	281.536	2.296[a]	7.584	36.057	89.758
T5–T42	T5	301.742	3.459	−6.057	24.578	81.522[c]
	T42	313.442	4.162	−6.332[b]	29.182	106.626
T35–T42	T35	259.697	0.876[a]	2.904	25.864	99.687
	T42	281.533	2.278	2.591[b]	26.393	99.427

[abc]The acoustics of the same tone in different pairs because they have different onsets and are influenced by the coarticulation of the onset.

7.3 Participants and procedures

The participants and the equipment in the current experiment are the same as the experiment in Chapter 6 (30 participants; women = 17, men = 13). The current experiment was conducted after the citation tone identification experiment in Chapter 6. The participants could take an optional ten-minute break between the two experiments. The experimenter is the same as for the previous experiment. All the instructions were delivered in Xiapu Min by the experimenter. The 40 tokens in the stimuli were presented to the participants once, resulting in 40 trials in total. The stimuli were divided into two blocks. The first block presented the recordings from Speaker #1 (20 tokens). The second block presented the recordings from Speaker #9 (20 tokens). The ordering of the stimuli within each block was randomized for each participant. Before the start of the test trials, four practice trials with the same task structure but different stimuli were given to the participants to help them familiarize with the task.

Listeners were told that there were compound words whose second words were masked by noise. They were asked to select the compound based on the recording of its first word. The options in each trial were two compounds whose first segments had the same segmental structure and were neutralized in tone (as shown in Table 7.2). Figure 7.3 shows what the participants see during a trial. In a given trial, the recording was played automatically in the beginning. Listeners were presented with the two target compounds on the screen, and selected which compound they heard from the two options. The options were displaced horizontally on the screen in Chinese characters. The ordering of the options was randomized for each trial and each participant. Listeners could click the "Replay" button on the screen and listen to the recording again as many times as desired. After they made a selection and were ready to move on to the next trial, they clicked the "Submit" button on the screen to proceed to the next trial.

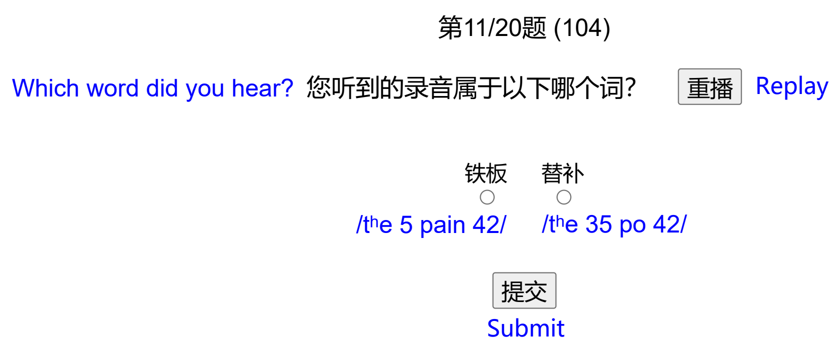


Figure 7.3. A sample page of the test trial for the compound identification task. Texts in blue are the translations of the content on the page. The translations were not present during the experiment.

The results from one participant are excluded from the analysis because they had self-identified hearing difficulties. Data from twenty-nine participants are included in the analysis. The participants were encouraged to report to the experimenter if they do not know a word in the choices of the identification task. One participant (Participant #7) did not know the word /ka 42 θe 42/ 假设 "hypothesis" in the stimuli. Thus, all the trials including that word as an option were excluded from analysis (8 tokens). Participant #12 reported that, in the trial where the stimuli is a token of /θi 5 ti 35/ 湿地 "wetland" produced by Speaker #1, the recording that they heard neither option on the screen. Their answer to that trial was excluded from the analysis. There are 1151 data points in total for analysis (40 tokens * 29 participants – 9 errors).

7.4 Results

The R script for the data analysis and the raw data are available in Supplementary Materials S8 and S9 at https://doi.org/10.17605/OSF.IO/3F5MN. If the participant chose the answer that matched the word of the original recording, the answer to that trial coded as correct. I calculated the accuracy of each contrast pair, and test whether it is significantly above, at, or below chance using binomial tests (41). The results are presented in Table 7.4. Only one contrast has above-chance accuracy (65%). All other four contrasts have an accuracy at chance level.

(41) Above chance or not:

$binom.test(number\ of\ sucess, number\ of\ trials, 0.5, ``greater")$

At chance or not:

$binom.test(number\ of\ sucess, number\ of\ trials, 0.5)$

Below chance or not:

$binom.test(number\ of\ sucess, number\ of\ trials, 0.5, ``less")$

Table 7.4. Identification accuracy of each neutralization pair

Neutralization	Correct	Incorrect	Correct rate
T2–T44	128	104	55.17% (At chance)
T23–T44	127	105	54.74% (At chance)
T5–T35	153	79	65.95% (Above chance)
T5–T42	118	113	51.08% (At chance)
T35–T42	111	113	49.55% (At chance)

In order to further investigate whether the contrasts that received at-chance accuracy are due to random guessing, or due to bias towards one specific tone category, I calculated the hit and false alarm rates of each tone in each tonal contrast based on Signal Detection Theory

(SDT) (Stanislaw and Todorov 1999; MacMillan 2002). Each tone in each contrast is used as the reference level when calculating the hit and false rates. In the identification task for the T2–T23 contrast, I first set T2 as the reference level, then defined underlying tone being T2 and response being T2 as a "hit"; underlying tone being T2 and response being T23 as a "miss"; underlying tone being T23 and response being T2 as a "false alarm"; underlying tone being T23 and response being T23 as a "correct rejection". Then, I set the reference level as T23, and define underlying tone being T23 and response being T23 as a "hit"; underlying tone being T23 and response being T2 as a "miss"; underlying tone being T2 and response being T23 as a "false alarm"; underlying tone being T2 and response being T2 as a "correct rejection". The hit rate and false alarm rate is calculated by formulars in (42). In addition, I calculated the d' of each neutralization pair using formular (43) (Stanislaw and Todorov 1999; MacMillan 2002). The d' represents the sensitivity of the differences between the options in identification task. A perfect discrimination (100% hit and 0% false alarm) has a d' of positive infinity. An at-chance guess has a d' close to zero (hit rate = false alarm rates). The d' is negative when the false alarm rate is higher than the hit rate, indicating that the listeners consistently mislabel Category A as Category B and vice versa. The percentage of each tonal response for each underlying tone for each contrast pair is presented in Figure 7.4. The count of the four SDT categories for each tone, along with the hit and false alarm rates, and d', are presented in Table 7.5. The hit rates are compared with the chance level (50%) using binomial tests (44).

(42) Hit rate: $hit/(hit + miss)$

False alarm rate: $false\ alarm/(false\ alarm + correct\ rejection)$

(43) $d' = z(hit\ rate) - z(false\ alarm\ rate)$

(44) Above chance or not:

$binom.test(hit, hit + miss, 0.5, ``greater")$

At chance or not:

$binom.test(hit, hit + miss, 0.5)$

Below chance or not:

$binom.test(hit, hit + miss, 0.5, ``less")$

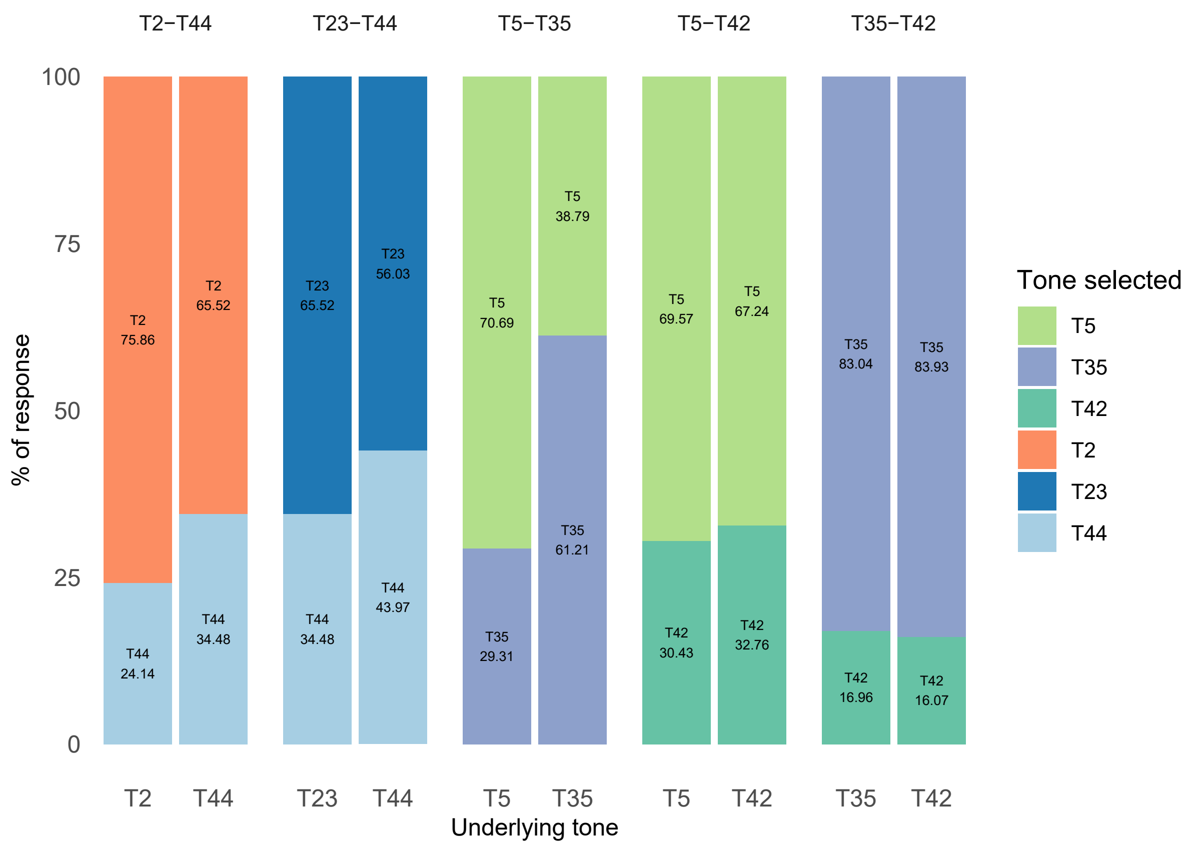


Figure 7.4. Responses of each tone in each neutralization pair

Table 7.5. Count of hit, false alarm, miss, and correctly reject for each tone in each contrast pair. The same set of responses within the same contrast pair is calculated twice, once with each underlying tone as the reference level. The d' of different reference level in the same contrast pair is the same. The content in the parenthesis after the hit rate indicates whether the hit rate is above, at, or below chance.

Neutralization	Underlying tone	Hit	False alarm	Miss	Correct rejection	Hit rate	False alarm rate	d'
T2–T44	T2	88	76	28	40	0.759 (Above)	0.655	0.303
	T44	40	28	76	88	0.345 (Below)		
T23-T44	T23	76	65	40	51	0.655 (Above)	0.560	0.241
	T44	51	40	65	76	0.440 (Chance)	0.345	
T5–T35	T5	82	45	34	71	0.707 (Above)	0.388	0.829
	T35	71	34	45	82	0.612 (Above)	0.293	
T5–T42	T5	80	78	35	38	0.696 (Above)	0.672	0.065
	T42	38	35	78	80	0.328 (Below)	0.304	
T35–T42	T35	93	94	19	18	0.830 (Above)	0.839	−0.036
	T42	18	19	94	93	0.161 (Below)	0.170	

As Table 7.5 shows, the d' for T5–T42 and T35–T42 are close to zero, suggesting the listeners are not sensitive to the difference between the neutralized tones in each pair. The d' is highest for T5–T35 pair, suggesting that listeners have relatively higher sensitivity towards the difference between T5 and T35 in the neutralized forms. Now let's looking into the hit rate and false rate of each tone, respectively. For the contrast T5–T35, the hit rate of T5 and

T35 is both above chance. The false alarm rate of both category is relatively small. It is in accordance with the overall identification accuracy of T5–T35 contrast is significantly above chance. In contrast, T2 in T2–T44, T23 in T23–T44, T5 in T5–T42, and T35 in T35–T42 all have an above-chance hit rate. Their false alarm rates are also larger than 50%. Moreover, the other tone in the category all receive at- or below-chance hit rate. This indicates that no matter what the underlying tone is, the listeners are more inclined to choose T2, T23, T5, and T35 over the other category in the contrast. In other words, the listeners have a bias towards T2, T23, T5, and T35 in those contrast pairs. The bias is very likely due to lexical frequency. During the informal conversations between the experimenter and the participants, several participants pointed out that the word /ka 42 θe 42/ 假设 "hypothesis" (T42 in T35–T42) is a less frequent word than /ka 35 kai 42/ 价格 "price" (T35 in T35–T42). One participant claimed that they did not know the pronunciation of /ka 42 θe 42/ 假设 "hypothesis", and all their answers for T35–T42 are excluded from analysis. The same could be true for T2 in T2–T44, T23 in T23–44, and T35 in T35–T42.

However, for neutralization pairs with a bias towards one category, it is still possible that the listeners are sensitive to the differences among two categories. Let's use T2–T44 as an example. It is possible that, although there is a bias towards T2 in general, the responses of T2 are significantly less when the underlying tone is T44 than when the underlying tone is T2. Such responses indicate that the listeners are sensitive to the underlying tone differences, despite the influence of lexical frequency. Thus, I perform statistical test with the response as the dependent variable, and the underlying tone as the independent variable (Model 45), and test within each neutralization contrast pair, whether the responses differ significantly by the underlying tone.

(45) $glmer(Response \sim Underlying\ tone + (1|subject), family = binomial(link = ``logit"))$

The results of Model (45) are in Table 7.6. The results show that the responses to the T2–T44 and the T5–T35 contrast pairs differ significantly by the underlying tone in each pair.

The underlying tone has a significant effect ($p < 0.05$) for T2–T44 and T5–T35 pairs. Less T2 responses were elicited when the underlying tone is T44 ($b = -0.868$, $p = .027$). Less T5 responses were elicited when the underlying tone is T35 ($b = -2.113$, $p < .001$). Combining these results with the descriptive statistics of the hit and false alarm rate, I conclude that the listeners are sensitive to the difference between T2 and T44, and between T5 and T35 in sandhi forms. Within the pair of T2–T44, there is bias towards identifying the neutralized tone as T2, probably due to the higher lexical frequency of T2.

Table 7.6. Statistics of Model (45).

		Estimate	Std. Error	z value	Pr(>|z|)	
T2–T44 (Reference level: Response T2 = 1 Underlying T2 = 1)	(Intercept)	2.047	0.566	3.619	< 0.001	
	Underlying T44	-0.868	0.393	-2.210	0.027	*
T23–T44 (Reference level: Response T23 = 1 Underlying T23 = 1)	(Intercept)	0.929	0.397	2.338	0.019	
	Underlying T44	-0.582	0.322	-1.809	0.071	
T5–T35 (Reference level: Response T5 = 1 Underlying T5 = 1)	(Intercept)	1.350	0.455	2.965	0.003	
	Underlying T35	-2.113	0.397	-5.320	< 0.001	***
T5–T42 (Reference level: Response T5 = 1 Underlying T5 = 1)	(Intercept)	2.673	1.109	2.410	0.016	
	Underlying T42	-0.281	0.467	-0.603	0.546	
T35–T42 (Reference level: Response T35 = 1 Underlying T35 = 1)	(Intercept)	1.785	0.336	5.309	< 0.001	
	Underlying T42	0.070	0.374	0.187	0.852	

The following question is, for the contrasts of T2–T44 and T5–T35, where listeners show differentiation between the neutralized tones, what acoustic cue do listeners use to identify the target syllable as one tone over the other? In order to answer this question, I regressed the responses of each contrast on the acoustic parameters of F0 (semi-tone), duration, H1*–H2*, and HNR. The duration of the stimuli is in a left-skewed distribution. In order to transform it into a normal distribution, I first subtracted each duration from the maximum duration of all stimuli plus 1, then take the square-root value of the difference (R code in 46). Then I transformed F0 (semi-tone), duration (after the square-root transformation), H1*–H2*, and HNR into z-scores to unify their scale. Last, I regressed the responses to the contrast pairs of T2–T44, and T5–T35 on the z-scored acoustic parameters. The R code of the regression is shown in (47). Because the data size within each contrast pair is relatively small, I performed model comparisons of the complex Model 47 with simpler models that drop one predictor at a time using ANOVA. If dropping one predictor does not make a significant difference to the model accuracy, I dropped that variable from Model (47).

(46) $Duration\ (transformed) = sqrt(max(Duration) + 1 - Duration)$

(47) $glmer(Response \sim F0 + Duration + H1^{*}{-}H2^{*} + HNR$
$+ (1|Participant), family = binomial(link = ``logit"))$

For the T2–T44 contrast, the ANOVA tests show that having H1*–H2* and HNR in the model as predictors do not improve the model accuracy significantly. Thus, H1*–H2* and HNR were dropped from the model. A simpler model with only F0 and Duration remains. The results show that lower F0 ($b = -1.308$, $p = .016$) and shorter duration ($b = 0.913$, $p = .010$) leads to a higher likelihood of having T2 as the response. Higher F0 and longer duration leads to a higher likelihood of T44 response.

For the T5–T35 contrast, the ANOVA tests show that having F0, H1*–H2*, and HNR in the model do not improve the model significantly. Thus, those three predictors are dropped, and a simpler model with Duration as the only predictor is used. The results show that shorter duration ($b = 1.364$, $p < .001$) leads to a higher likelihood of T5 response. Longer duration leads to a higher likelihood of T35 response.

7.5 Summary

Using natural stimuli of T2–T44, T23–T44, T5–T35, T5–T42, and T35–T42, I find that listeners of Xiapu Min are sensitive to the acoustic differences between T2 and T44 and between T5 and T35. As predicted, checked tones – T2 and T5 – are elicited by shorter duration. Lower F0 also elicits more T2 responses. Listeners' identification results do not significantly differ by the underlying tone in the contrast of T23–T44, T5–T42, or T35–T42, indicating that they are not able to differentiate the natural stimuli of T23 vs. T44, T5 vs. T42, or T35 vs. T42. There is bias towards one specific tonal category in the contrasts of T2–T44, T23–T44, T5–T42, and T35–T42, probably due to an imbalanced word frequency between the two compounds under contrast.

The fact that short duration elicits checked tone responses despite the loss of glottalization in checked tones in sandhi forms confirms my hypothesis that the short duration is independent of the glottal coda. It is also in accordance with the results of the perception experiment in Chapter 6, which show that duration a more effective factor than glottalization in eliciting citation checked tones.

Now we can revisit the question posited at the beginning of this chapter: which type of neutralization, as proposed by Dinnsen 1985, do Xiapu Min neutralized sandhi tones belong to? The categorization differs by contrast. As the results from the production study in Chapter 5 Section 5.2 shows, T23–T44 and T35–T42 are not differentiated on the acoustic space in the production. The identification test shows that these two contrasts are not differentiated by the listeners in perception either. They belong to Type A: complete neutralization in production and perception. Contrast T5–T42 are well-differentiated in production. However, the listeners are not sensitive to the differences between T5 and T42. Therefore, T5–T42 should belong to Type B: incomplete neutralization in production and complete neutralization in perception. Contrast T2–T44 also has differences in production, and the listeners are able to differentiate T2–T44 in perception. However, the cue that the listeners use in perception differs from the acoustic differences found in the production. The listeners chose tokens with lower F0 as T2. However, in the stimuli used in the current identification task, T2 has higher mean F0 than T44, and the difference is small (T2: 200 Hz, 3.76 semitones below the baseline; T44: 189 Hz, 4.57 semitones below the baseline.) The production results in Figure 5.2 based on the data from ten speakers also show that, in sandhi forms, T2 and T44 largely overlap in F0, and T2 has a higher mean F0 than T44. Thus, the listeners are using F0 cues that are not present in the natural production to differentiate T2 from T44 in perception. The possible motivation is that T2 has a low-falling F0 contour whereas T44 has a mid-level F0 contour in citation forms. The listeners might be appealing to the F0 property of the citation form of T2. On the other hand, short duration does differentiate T2 from T44 in production, and is also used as another perception cue. Thus, T2–T44 contrast belongs to two types – Type C: incomplete neutralization in production and perception (regarding the duration cue); and Type D: no differences in production but differences in perception (regarding the F0 cue). T5–T35 are primarily differentiated by duration in both production and perception. It thus belongs to Type C: differences in both production and perception. The production and perception results of the tonal neutralization in sandhi forms in Xiapu Min indicates that neutralization phenomena resulted from the same phonological process (i.e. tone sandhi) can belong to different types of neutralization. The results are summarized in Table 7.7.

Table 7.7. Typology of tonal neutralization in Xiapu Min

Type	Production differences	Perceptual differences	Examples
A	No	No	T23–T44, T35–T42
B	Yes	No	T5–T42
C	Yes	Yes	T2–T44 (duration), T5–T35 (duration)
D	No	Yes	T2–T44 (F0 difference)

The limitation of the current study is that the lexical frequency is not controlled in the contrast pair. This limitation is due to the lack of documentation and large-scale corpora of the language. In addition, each contrast is tested by a single pair of compounds. For contrasts such as T5–T42, where production differences are found, but the perception differences are not, the reason could be attributed to the influence of large difference in lexical frequency between the contrasts, and a lack of variation in the stimuli. In the future, a self-evaluated word-frequency test can be conducted before the identification experiment, so that we can select words of comparable lexical frequencies. Multiple minimal pairs can be tested for the same tonal contrast to increase the representativeness of the results.

The other direction for the future study is using (re)synthesized stimuli to control the variation in the stimuli, as the perception study for the citation tone did in Chapter 6. For example, for T2–T44 contrast, where F0 differences were not found in the production, but was correlated to the listeners' categorization of the tone in perception, we can vary the F0 into multiple levels with all other acoustic parameters controlled. The results will help verify whether it is the case that listeners' responses vary by F0 systematically. For T5–T42, which has a below-chance categorization accuracy, we can increase the degree of duration variation, and see whether more extreme short and long durations will elicit different categorization results.

Chapter 7, in full, is currently being prepared for submission for publication of the material. Chai, Yuan, and Shihong Ye. "Acoustic cues for checked tone perception in Xiapu Min". The dissertation author was the primary investigator and author of this material.

Chapter 8

Conclusion

8.1 Definition of checked

The first research question of this dissertation is: at the phonological level, when is it necessary to posit a phonological constituent of "checked"? The answer, based on the survey of Zapotec and Chinese languages, is that there are two circumstances: one is when the language has late-phased glottalization as a phonation type (i.e., checked phonation); the other is when there are checked syllables coocurring with checked tones. The criterion that I use in this dissertation is: if there is an opposition between the checked consituents and the unchecked ones, and there is no existing phonological constituent that can account for such an opposition, it is necessary to identify a checked category in the phonology of that language.

8.2 What counts as a checked phonation type?

Checked phonation is a phonological category. The phonological structure of checked phonation can be reduced to the phonological feature of [+constricted glottis] and a temporal specification of late-phased alignment with the vowel that bears the [+constricted glottis] feature. The criteria of checked phonation are that 1) there is late-phased glottalization in the vowels (Vʔ); 2) the glottal stop in Vʔ is suprasegmental. The checked phonation can be allophonic to the rearticulated phonation (e.g., Tilquiapan Zapotec), or be contrastive with another glottalized phonation – rearticulated phonation ($V^{ʔ}V$) (e.g., Isthmus Zapotec). It is necessary to posit a phonological category of "checked" for /Vˀ/, because there is no existing phonological categories that account for phonation types differing in timing of glottalization.

8.3 What counts as a checked syllable/tone?

Checked syllable/tone is a descriptor of the phonotactic constraint between certain closed syllables and specific tones. The phonological structure of this kind of "checked" can be reduced to the combination of two phonological categories – syllables closed by stops and tone. For a syllable/tone to be considered checked, the criteria are: 1) the language has syllables ending in stops; 2) stop-closed syllables bear a different set of tones from open or sonorant-closed syllables; 3) if the only stop coda in the language is a glottal stop, the glottal stop is a segment. Checked syllable and tone always occur together. There is no existing term accounting for such phonological restriction between syllable structure and tone. Many Chinese languages (e.g. Taiwanese Min, Shanghainese, Hakka), and neighboring languages such as Vietnamese and Burmese have such kinds of checked syllables and tones.

When identifying whether a constituent has checked phonation, or checked syllable and checked tone, one important criterion is determining whether the glottal stop in Vʔ and/or VʔV should be treated as a segment or not. I have reviewed criteria from various studies for determining glottal stop as a segment or not, and proposed a general principle based on phonological evidence: if Vʔ syllables behave the same as other closed syllables in phonological processes, the glottal stop is a segment; if Vʔ syllables behave the same as open syllables, the glottal stop is a suprasegment. I summarize the criteria for Zapotec and Chinese languages as following:

- Zapotec languages
  - Phonetic glottal stop analyzed as suprasegmental – /Vˀ/
    - [Vʔ] patterns with open syllables in phonological processes: Texmelucan Zapotec suffixation;

    * [Vʔ] patterns with non-modal phonation types and is in opposition to modal phonation in phonological processes: Guienagati Zapotec coda distribution.
  - Phonetic glottal stop analyzed as segmental – /V+ʔ/
    * Vʔ has same stress distribution as VC but different from V: Cajonos Zapotec;
    * Vʔ cannot be followed by consonant coda but V and VʔV can: Cajonos Zapotec.
  - Phonetic glottal stop analyzed as suprasegmental /V^{ʔ}V/
    * [VʔV] exhibits the phonological properties of monosyllables, such as having the same size of tonal inventory; having the same size of vowel inventory as monosyllables: Quiaviní Zapotec;
    * Native speakers judge [VʔV] as one single syllable in activities such as tapping, humming, clapping, or whistling per syllable: Yalálag Zapotec.

- Chinese languages
  - Phonetic glottal stop analyzed as segmental – /V+ʔ/
    * [Vʔ] shares the same vowel quality as [VN] and differs from [V]: Shaoxing Wu;
    * [Vʔ] prevents onset assimilation in /Vʔ +CV/, whereas [V] licenses onset assimilation: Putian Jiangkou Min;
    * The underlying form of [ʔ] in [Vʔ] is an oral consonant: Nantong Jianghuai Mandarin;
    * [Vʔ (<*-k)] prevents onset assimilation in [Vʔ (<*-k) + CV], whereas [V] licenses onset assimilation in [V + CV]. The glottal stop in [Vʔ (<*-k)] is segmental: Fuzhou Min
  - Phonetic glottal stop analyzed as suprasegmental – /V^{ʔ}/
    * [Vʔ (<* -ʔ)] licenses onset assimilation in [Vʔ + CV], whereas [Vʔ (<*-k)] prevents onset assimilation in [Vʔ (<*-k) + CV]. The glottal stop in [Vʔ (<*-ʔ)] is suprasegmental: Fuzhou Min.

8.4 The commonality between checked phonation and checked syllables/tones

As their names imply, checked phonation and checked syllables/tones refer to different phonological structures. Yet they are both called "checked" because of their shared phonetic nature. They both have abrupt offset of voicing, reinforced by either late-phased glottalization or obstruent coda. The purpose of the current dissertation is to provide a criterion of determining, when we observe abrupt offset of voicing in a language, whether such phonetic phenomenon should be regarded as a distinct phonological constituent. The principle of this criteria is – whether there is existing phonological feature, category, or constraint that can account for the phonological constituent where the abrupt offset of voicing is observed. Phonation with late-phased glottalization and closed syllables with a restricted tonal distribution are frequently found to have an abrupt offset of voicing. And as explained in Sections 8.2 and 8.3, neither phonation with late-phased glottalized nor the constraint between closed syllable and tone can be accounted by an existing phonological category or constraint. Thus, we propose that those phonetically checked constituents are distinct phonological category or constraint.

8.5 Phonetics of checked syllable and checked tone – the case study of Xiapu Min

Besides discussing the phonological properties of checked cross-linguistically, I have also provided empirical evidence of the phonetic properties of Xiapu Min, a language that has checked syllable and checked tone. The studies have provided the first quantitative analysis of the tone system of this language. Xiapu Min has high and low checked tones associated with Vʔ checked syllables. In citation forms, high checked tone has high-falling F0; low checked tone has low-falling F0. Both checked tones are glottalized at the second half of the vowel, and have shorter duration than unchecked tones. No difference in vowel quality is found between checked and unchecked tones. In sandhi forms, checked tones are phonologically neutralized with unchecked tones into a mid-level or high-level tone. Yet here I found at the phonetic level, checked tones can still be differentiated from unchecked tones, primarily by duration. Checked tones continue to be shorter than unchecked tones even after neutralization. This indicates that even when the glottal constriction in the checked syllables are lost in sandhi forms, the short duration is still preserved. In Xiapu Min, short duration is not a by-product of the glottal coda constriction; instead, it is a phonetic target of checked syllables in Xiapu Min.

Perception studies of Xiapu Min checked tones in citation forms and sandhi forms illustrate which acoustic cues are essential for perceiving checked tones. The results show that in citation forms, F0, duration, and glottalization are all effective cues for eliciting checked tones. F0 and duration are more important cues than glottalization. In sandhi forms, two pairs of neutralization (T2–T44 and T5–T35) are still distinguished by the listeners, despite the purported phonological neutralization between checked and unchecked tones. This indicates that the neutralization of checked and unchecked tones is phonetically incomplete in perception as well. The primary cue that listeners use to distinguish checked tone from unchecked tone is duration and F0. The results of the phonetic studies reflect the correspondence between the production and perception of checked syllables and tones in Xiapu Min. Cues that distinguish checked syllables from unchecked ones in production are also essential cues for listeners to determine whether a syllable is checked. The results from the perception studies for checked tones in both citation and sandhi form further confirm the independence between duration and glottalization in the production study. Glottalization is a less important cue than duration for perceiving checked tone in citation forms, and has not been found as an effective cue for perceiving checked tone in sandhi form. Duration, instead, can elicit checked tones independently of glottalization in both citation and sandhi forms.

8.6 Future directions

I propose two directions for future studies on the phonetics and phonology of checked constituents. First, future studies can expand the survey to more languages involve phonetically checked constituents, and explore whether there are different types of checked constituents at the phonological level. The expanded survey can start from neighboring language families of Zapotec and Chinese languages, such as Chatino (Popoloca-Zapotecan), Mixtecan (Amuzgo-Mixtecan), Chinantecan (Western Otomanguean), Lolo-Burmese (Sino-Tibetan), Hmong-Mien, Austroasiatic languages, among others.

Second, future studies can study the phonetic properties of languages with checked and rearticulated phonation types in more detail. Checked and rearticulated phonations are both types of glottalized phonation, but differ in the phasing of the glottalization relative to the vowels. As discussed in Section 2.2, checked and rearticulated can be contrastive phonation types in a language. Thus, it is meaningful to comprehensively describe the phonetic nature of those two phonation types that are phonologically in opposition. Specifically, future studies should examine how the phasing of glottalization matters in production using acoustic and articulatory tools, and whether they have phonetic differences in other dimensions (e.g., F0, duration, vowel quality). In terms of perception, future studies should study what cue(s) listeners use to differentiate checked and rearticulated phonation types. One cue of focus is the timing of glottalization. Because the timing of glottalization is contrastive for languages with checked phonation and rearticulated phonation, changing the timing of glottalization to the beginning, middle, end, or the entirety of the vowel will likely elicit identifications of different phonation types. Although no phonation type in Zapotec languages have been found to be characterized by early-phased glottalization on the vowels, I hypothesize that putting glottalization at the beginning of the vowel is more likely to elicit a percept of rearticulated phonation, since early-phased glottalization is closer to mid-phased glottalization (found in rearticulated phonation) than late-phased glottalization (found in checked phonation) in terms of the timing of glottalization. In addition, future studies should also test whether cues other than the timing of glottalization, but that also have been found to differ between checked and rearticulated phonations in production (e.g., F0, duration, vowel quality), are also used by listeners to perceive the differences between checked phonation and rearticulated phonation. Currently, I am working with speakers of Yateé Zapotec (Northern Core, yate1242, zty), a Zapotec language that have contrastive checked and rearticulated phonations (Jaeger and Van Valin 1982). I will investigate the above research questions using data from Yateé Zapotec in the future. With more descriptive data of the phonetic properties of checked phonation, we can compare the phonetics of checked phonation with the phonetics of checked syllables and tones, and investigate whether the two checked constituents that differ in phonological structure also differ in certain phonetic properties.

The methodology and observations in this dissertation can be expanded beyond the phenomenon of checked constituents. For any phonetic realization, if we want to determine whether it is necessary to assign a novel phonological feature to it, the essential question is whether it contains features that cannot be explained by the existing phonological features. The revised typology of tone–phonation interaction can be used to analyze the relation between tone and any other phonation in any language. The survey of **Rù* syllables/tones in Middle Chinese, and checked syllables/tones in Modern Chinese, can be used to study the sound change pattern of checked syllables and tones in Chinese. The loss of checked syllables has three major steps: debuccalization of the oral codas, lengthening of the vowel, and loss of the glottal coda. The paths can be different among languages due to the different sequencing of these three steps. We can compare across different languages from different sub-families, in order to determine for a given language, which stage of sound change it is currently at. We can make predictions of what its next stage of sound change will be by referring to languages that are further down in the path of the sound change. We can also make reconstructions of the language to suggest proto-forms by referring to languages that are the earlier stage of the sound change path.

Bibliography

Ali, Jehad, Rehanullah Khan, Nasir Ahmad, and Imran Maqsood. 2012. "Random forests and decision trees." *International Journal of Computer Science Issues* 9 (5): 272–278.

Andruski, Jean E., and Martha Ratliff. 2000. "Phonation types in production of phonological tone: the case of Green Mong." *Journal of the International Phonetic Association* 30:37–61. https://doi.org/10.1017/S0025100300006654.

Ao, Benjamin Xiaoping. 1993. "Phonetics and phonology of Nantong Chinese." PhD diss., Ohio State University. Accessed August 23, 2022. https://etd.ohiolink.edu/apexprod/rws_olink/r/1501/10?clear=10&p10_accession_num=osu1487844485897534.

Audacity team. 2022. *Audacity (Version 3.1.3) [Computer program].* Accessed June 25, 2022. http://audacityteam.org/.

Avelino, Heriberto. 2004. "Topics in Yalálag Zapotec, with particular reference to its phonetic structures." Dissertation, University of California, Los Angeles.

———. 2016. "Phonetics in Phonology: A Cross-Linguistic Study of Laryngeal Contrast." In *The Phonetics and Phonology of Laryngeal Features in Native American Languages,* edited by Heriberto Avelino, Matt Coler, and W. Leo Wetzels, 157–179. Leiden/Boston: Brill.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using **lme4**." *Journal of Statistical Software* 67 (1). https://doi.org/10.18637/jss.v067.i01.

Bauer, Robert S., and Stephen Matthews. 2017. "Cantonese." In *The Sino-Tibetan Languages,* Second edition, edited by Graham Thurgood and Randy J. LaPolla, 169–184. Routledge language family series. London; New York: Routledge.

Beam de Azcona, Rosemary Grace. 2004. "A Coatlan-Loxicha Zapotec grammar (Mexico)." PhD diss., University of California, Berkeley.

Benn, Joshua. 2016. "Consonant-Tone-Phonation Interactions in Guienagati Zapotec." In *5th International Symposium on Tonal Aspects of Languages (TAL 2016),* 125–128. Buffalo, New York: ISCA, May. https://doi.org/10.21437/TAL.2016-27.

Benn, Joshua. 2021. “The phonetics, phonology, and historical development of Guienagati Zapotec.” PhD diss., State University of New York at Buffalo.

Bennett, Ryan. 2016. “Mayan phonology: Mayan phonology.” *Language and Linguistics Compass* 10, no. 10 (October): 469–514. https://doi.org/10.1111/lnc3.12148.

Bishop, Jason B. 2007. “Incomplete neutralization in Eastern Andalusian Spanish: Perceptual consequences of durational differences involved in s-aspiration.” In *Proceedings of the 16th International Congress of Phonetic Sciences,* 1765–1768. Saarbrücken, Germany.

Boersma, Paul, and David Weenink. 2021. *Praat: doing phonetics by computer [Computer program],* November. Accessed April 30, 2022. http://www.praat.org/.

Borroff, Marianne L. 2007. “A landmark underspecification account of the patterning of glottal stop.” PhD diss., Stony Brook University.

Braver, Aaron. 2014. “Imperceptible incomplete neutralization: Production, non-identifiability, and non-discriminability in American English flapping.” *Lingua* 152 (December): 24–44. https://doi.org/10.1016/j.lingua.2014.09.004.

Brockmann, Meike, Michael J. Drinnan, Claudio Storck, and Paul N. Carding. 2011. “Reliable jitter and shimmer measurements in voice clinics: the relevance of vowel, gender, vocal intensity, and fundamental frequency effects in a typical task.” *Journal of Voice* 25:44–53.

Browman, Catherine P., and Louis M. Goldstein. 1986. “Towards an articulatory phonology.” *Phonology Yearbook* 3:219–252. https://doi.org/10.1017/S0952675700000658.

Brunelle, Marc. 2009. “Tone perception in Northern and Southern Vietnamese.” *Journal of Phonetics* 37, no. 1 (January): 79–96. https://doi.org/10.1016/j.wocn.2008.09.003.

Brunelle, Marc, and Joshua Finkeldey. 2011. “Tone perception in Sgaw Karen.” In *Proceedings of the ICPhS XVII 2011,* 372–375. Hong Kong.

Bueno Holle, Juan Jose. 2016. “Information structure in Isthmus Zapotec - ProQuest.” PhD diss., The University of Chicago.

Campbell, Lyle. 1974. “Theoretical Implications of Kekchi Phonology.” *International Journal of American Linguistics* 40 (4): 269–278.

Cao, Zhiyun 曹志耘. 2002. “Wuhuiyu rusheng yanbiande 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,* no. 5, 441–446. Accessed August 24, 2022. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2002&filename=YWZG200205006&uniplatform=NZKPT&v=s0B9R-KhbM2fGmUgWpfGAVRwlZ2h2Jr5IcZir7Uix_vSZFISSDNnpfFslQ37A_qV.

Chan, Marjorie K. M. 1987. “Tone and Melody in Cantonese.” *Annual Meeting of the Berkeley Linguistics Society* 13 (September): 26. https://doi.org/10.3765/bls.v13i0.1828.

———. 1993. *Fuzhou phonology: A non-linear analysis of tone and stress.* Ann Arbor, MI: A Bell & Howell Company.

———. 1997. “Fuzhou glottal stop: Floating segment or correlation of close contact?” In *Studies in Chinese phonology,* edited by Jialing Wang and Norval Smith, 275–289. Berlin: Mouton de Gruyter. ISBN: 3-11-013953-7.

Chao, Yuen Ren. 1930. “A system of “tone-letters”.” *Le Maître Phonétique Troisième Série* 8 (45): 24–27.

Chappell, Hilary, and Christine Lamarre. 2005. *A Grammar And Lexicon of Hakka.* Vol. 8. Ecole des Hautes Etudes en Sciences Sociales. ISBN: 978-2-910216-07-8, accessed January 15, 2022. https://hal-inalco.archives-ouvertes.fr/hal-01386151.

Chávez-Peón, Mario E. 2010. “The interaction of metrical structure, tone, and phonation types in Quiaviní Zapotec.” PhD diss., University of British Columbia.

———. 2011. “Non-modal phonation in Quiaviní Zapotec: An acoustic investigation.” In *Memorias del V Congreso de Idiomas Indígenas de Latinoamérica,* 1–24. Universidad de Texas en Austin. Accessed August 24, 2022. https://ailla.utexas.org/sites/default/files/documents/Chavez_CILLA_V.pdf.

Chen, Matthew. 1970. “Vowel Length Variation as a Function of the Voicing of the Consonant Environment.” *Phonetica* 22 (3): 129–159. https://doi.org/10.1159/000259312.

———. 1976. “From Middle Chinese to Modern Peking.” *Journal of Chinese Linguistics* 4 (2/3): 113–277. Accessed May 24, 2022. https://www.jstor.org/stable/23752895.

Chen, Qin 陈勤. 2006. “Gutian Pinghu hua shengmu yuliu yinbian 古田平湖话声母语流音变 [The allophonic variation of Pinghu speech’s initials].” *Journal of Minjiang University (Natural Sciences)* 27 (6): 93–97. Accessed August 24, 2022. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2006&filename=FSXB200606022&uniplatform=NZKPT&v=NgF1nVq6yfis12tbDR_6tozaJhRnIISUKR5GmuC_3HBy9wwNxVc4mF_tNAmlM1pL.

Chen, Si, and Caroline R. Wiltshire. 2013. “Tone realization in younger versus older speakers of Nanjing dialect.” In *Studies in Chinese Language and Discourse,* edited by Zhuo Jing-Schmidt, 2:147–170. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/scld.2.07che.

Chen, Xueping 陈雪萍. 2017. "Putian Daitou fangyan yuyin yanjiu 莆田埭头方言语音研究 [The phonetic study on Putian Daitou dialect]." Master's thesis, Xiamen University. Accessed August 24, 2022. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CMFD&dbname=CMFD201901&filename=1017105351.nh&uniplatform=NZKPT&v=9ZHKkz0QfG5Mv62zygiRK0uXg7gq8xJWyDq7iYbfOxo7-GkgabrEoO--W1UKthNC.

Chen, Yiya, and Carlos Gussenhoven. 2015. "Shanghai Chinese." *Journal of the International Phonetic Association* 45, no. 3 (December): 321–337. https://doi.org/10.1017/S0025100315000043.

Cheng, Teresa M. 1973. "The phonology of Taishan." *Journal of Chinese Linguistics* 1 (2): 256–322. ISSN: 0091-3723, accessed January 16, 2022. https://www.jstor.org/stable/23749797.

Cheung, Kwan-hin. 1986. "The phonology of present-day Cantonese." PhD diss., University College London. Accessed August 24, 2022. https://discovery.ucl.ac.uk/id/eprint/1317605/.

Chien, Yu-Fu, and Allard Jongman. 2019. "Tonal Neutralization of Taiwanese Checked and Smooth Syllables: An Acoustic Study." *Language and Speech* 62, no. 3 (September): 452–474. https://doi.org/10.1177/0023830918785663.

Córdova, Jennifer Lynn Heise. 2009. "Participant reference and tracking in San Francisco Ozolotepec Zapotec," 14. SIL e-Books 14. Dallas: SIL International. Accessed August 24, 2022. https://www.sil.org/resources/publications/entry/9269.

Crowhurst, Megan J., Niamh E. Kelly, and Amador Teodocio Olivares. 2016. "The influence of vowel laryngealisation and duration on the rhythmic grouping preferences of Zapotec speakers." *Journal of Phonetics* 58:48–70. https://doi.org/10.1016/j.wocn.2016.06.001.

De Maesschalck, R., D. Jouan-Rimbaud, and D.L. Massart. 2000. "The Mahalanobis distance." *Chemometrics and Intelligent Laboratory Systems* 50, no. 1 (January): 1–18. https://doi.org/10.1016/S0169-7439(99)00047-7.

Deng, Xiangzhang 邓享璋. 2007. "Minbei Minzhong fangyan yanjiu 闽北、闽中方言语音研究 [A phonetic study of the Northern Min dialect and Central Min dialect]." PhD diss., Xiamen University. Accessed August 24, 2022. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CDFD&dbname=CDFD9908&filename=2008058225.nh&uniplatform=NZKPT&v=ui-mJmYL2_OozR9p-2dMx-q9PQWZuSGUSae7pkEZhts-iv4P8eDZiLi9jy8fOyCS.

Dinnsen, Daniel A. 1985. "A re-examination of phonological neutralization." *Journal of Linguistics* 21 (2): 265–279. https://doi.org/10.1017/S0022226700010276.

Dong, Hongyuan. 2020. *A History of the Chinese Language.* 2nd ed. Second edition. | Abingdon, Oxon ; New York, NY : Routledge, 2021.: Routledge, December. https://doi.org/10.4324/9780429264665.

Du, Xiaoping 杜晓萍. 2021. “Fujian Shaowu Hongdun fangyan de shusheng cuhua 福建邵武洪墩方言的舒声促化 [On the pronunciation evolution from MC Non-Entering tones into Entering tone in Modern Hongdun dialect in Shaowu city, Fujian province].” *Dialect 方言* 2:214–222. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2021&filename=FYZA202102014&uniplatform=NZKPT&v=r50a5FZZJxZ_NXT22XNnGWFRFJIDxTuAEGn1fVv3qdFdHfLwyRWcp7gg_Q3jCz0w.

Duan, Wenjun, and Yuan Jia. 2015. “Contrastive study of focus phonetic realization between Jinan dialect and Taiyuan dialect.” In *2015 International Conference Oriental COCOSDA held jointly with 2015 Conference on Asian Spoken Language Research and Evaluation (O-COCOSDA/CASLRE),* 47–52. Shanghai, China: IEEE, October. https://doi.org/10.1109/ICSDA.2015.7357863.

Duanmu, San. 1994. “The phonology of the glottal stop in Garo.” *Linguistics of the Tibeto-Burman Area* 17 (2): 69–82. Accessed August 24, 2022. http://www-personal.umich.edu/~duanmu/Garo94.pdf.

Earl, Robert. 2011. “Gramática del Zapoteco de Tabaa.” April. Accessed August 24, 2022. https://www.sil.org/resources/archives/40770.

Esposito, Christina M. 2003. “Santa Ana Del Valle Zapotec Phonation.” Master’s thesis, University of California, Los Angeles.

———. 2010. “Variation in contrastive phonation in Santa Ana Del Valle Zapotec.” *Journal of the International Phonetic Association* 40 (2): 181–198. https://doi.org/10.1017/S0025100310000046.

———. 2012. “An acoustic and electroglottographic study of White Hmong tone and phonation.” *Journal of Phonetics* 40 (3): 466–476. https://doi.org/10.1016/j.wocn.2012.02.007.

Farnetani, Edda, and Shiro Kori. 1986. “Effects of syllable and word structure on segmental durations in spoken Italian.” *Speech Communication* 5, no. 1 (March): 17–34. https://doi.org/10.1016/0167-6393(86)90027-0.

Feng, Faqiang 冯法强. 2011. “Zhonggu 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. Accessed February 15, 2022. https://kns.cnki.net/kcms/detail/detail.aspx?filename=1011161190.nh&dbcode=CMFD&dbname=CMFD2011&v=OKLZRs3sXo2jbu39k1Ggnx6lORKrVJu7ifxAKRC9pVhqYACyA7RNF3dgYdHrZPdf.

Ferlus, Michel. 2009. “What were the four divisions of Middle Chinese?” *Diachronica* 26, no. 2 (July): 184–213. https://doi.org/10.1075/dia.26.2.02fer.

Frazier, Melissa. 2009. "The production and perception of pitch and glottalization in Yucatec Maya." PhD diss., The University of North Carolina at Chapel Hill University Libraries. https://doi.org/10.17615/SD6J-0526.

———. 2013. "The phonetics of Yucatec Maya and the typology of laryngeal complexity." *STUF - Language Typology and Universals* 66, no. 1 (April): 7–21. https://doi.org/10.1524/stuf.2013.0002.

Gafos, Adamantios I. 2002. "A Grammar of Gestural Coordination." *Natural Language and Linguistic Theory* 20 (2): 269–337. https://doi.org/10.1023/A:1014942312445.

Garellek, Marc. 2013. "Production and perception of glottal stops." PhD diss., University of California, Los Angeles.

———. 2019. "The phonetics of voice." In *Routledge Handbook of Phonetics,* edited by William Katz and Peter Assmann, 75–106. Oxford: Routledge.

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,* 1–28. https://doi.org/10.1017/S0025100321000116.

Garellek, Marc, and Christina M. Esposito. 2021. "Phonetics of White Hmong vowel and tonal contrasts." *Journal of the International Phonetic Association,* 1–20. https://doi.org/10.1017/S0025100321000104.

Garellek, Marc, and Patricia Keating. 2011. "The acoustic consequences of phonation and tone interactions in Jalapa Mazatec." *Journal of the International Phonetic Association* 41 (2): 185–205. https://doi.org/10.1017/S0025100311000193.

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, no. 2 (February): 1078–1089. https://doi.org/10.1121/1.4773259.

Garellek, Marc, and Marija Tabain. 2020. "Tongan." *Journal of the International Phonetic Association* 50 (3): 406–416. https://doi.org/10.1017/S0025100318000397.

Gittlen, Laura, and Stephen A. Marlett. 1985. "Ñumí Mixtec syllable structure and morphology." *Work Papers of the Summer Institute of Linguistics, University of North Dakota Session* 29 (1): 175–194. Accessed August 24, 2022. https://commons.und.edu/cgi/viewcontent.cgi?article=1277&context=sil-work-papers.

Green, Antony Dubach. 2015. "Word, foot and syllable structure in Burmese Syntax." In *Studies in Burmese Linguistics,* edited by Justin Watkins, 1–26. Canberra: Pacific Linguistics.

Gruber, James Frederick. 2011. "An Articulatory, Acoustic, and Auditory Study of Burmese Tone." PhD diss., Georgetown University.

Gu, Qian 顾黔. 2015. “Nanjing fangyan Dang Jiang liang rushengyunde gongshi bianyi 南京方言宕江两摄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 (4): 384–393. Accessed August 24, 2022. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2015&filename=YYKE201504006&uniplatform=NZKPT&v=VdwtPmg3bNUB8Cgb2JTkqMMu9PPQ3bisVdBU8dcZLcO5MIljMIabNruVXuKOx_T5.

Haudricourt, André G. 1954. “De l’origine des tons en vietnamien [The origin of tones in Vietnamese].” *Journal Asiatique* 242:69–82.

Hayes, Bruce. 2011. *Introductory Phonology.* OCLC: 1020573489. Oxford: Wiley-Blackwell. ISBN: 978-1-4443-6013-4.

Heiberger, Vicki L., and Yoshiyuki Horii. 1982. “Jitter and shimmer in sustained phonation.” In *Speech and Language,* 7:299–332. Elsevier. https://doi.org/10.1016/B978-0-12-608607-2.50016-9.

Heise, Jennifer Lynn. 2003. “Participant reference and tracking in San Francisco Ozolotepec Zapotec.” Master’s thesis, Graduate Institute of Applied Linguistics. Accessed August 24, 2022. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.531.5149&rep=rep1&type=pdf.

Hu, Na, Yuan Jia, and Bin Liu. 2012. “Phonetic and phonological realization of narrow focus in English declarative sentences by Zhenjiang EFL learners.” In *Proceedings of Speech Prosody 2012,* 394–397. Shanghai, China. Accessed August 24, 2022. https://www.isca-speech.org/archive_v0/sp2012/sp12_394.html.

Hu, Xiaomei 胡晓美. 2020. “Mindong Pingxihua yinyun cihui yanjiu 闽东平溪话音韵词汇研究 [A study of Pingxi dialect in eastern Fujian].” Master’s thesis, Shanghai Normal University. Accessed August 24, 2022. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CMFD&dbname=CMFD202002&filename=1020809180.nh&uniplatform=NZKPT&v=gn3NACdqsmPnTG6a-eGbv4FnTzX-bWRC3YBjFDWlF6KfkCwVuKFLPBCWEiZGn7DV.

Huang, Biyu 黄碧玉. 2013. “Putian Jiangkou fangyan yinxi yanjiu 莆田江口方言音系研究 [Resesarch on the phonology of Jiangkou Putian dialect].” Master’s thesis, Fujian Normal University. Accessed August 24, 2022. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CMFD&dbname=CMFD201402&filename=1013306897.nh&uniplatform=NZKPT&v=Dp8a2nCW4Ih6u8syDtb2zgX1I-Luz2P1ab1pqbGQ-_ujsAWlc6RwCK9ultxXt6Vp.

Huang, Jingyuan, and Lori L. Holt. 2009. “General perceptual contributions to lexical tone normalization.” *The Journal of the Acoustical Society of America* 125 (6): 3983–3994. https://doi.org/10.1121/1.3125342.

Huang, Yaqian. 2020. "Different attributes of creaky voice distinctly affect Mandarin tonal perception." *The Journal of the Acoustical Society of America* 147 (3): 1441–1458. https://doi.org/10.1121/10.0000721.

Huang, Yishan. 2018. "Tones in Zhangzhou: Pitch and Beyond." Publisher: The Australian National University. PhD diss., Australian National University.

Huffman, Marie K. 1987. "Measures of phonation type in Hmong." *The Journal of the Acoustical Society of America* 81 (2): 495–504. https://doi.org/10.1121/1.394915.

Hwa-Froelich, Deborah, Barbara W. Hodson, and Harold T. Edwards. 2002. "Characteristics of Vietnamese Phonology." *American Journal of Speech-Language Pathology* 11 (3): 264–273. https://doi.org/10.1044/1058-0360(2002/031).

Institute of Linguistics, Chinese Academy of Social Science. 2018. *Fangyan diaocha zibiao 方言调查字表 [Dictionary for surveys on dialects].* 1st ed. Beijing: The Commercial Press 商务印书馆.

Izenman, Alan Julian. 2013. "Linear discriminant analysis." In *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning,* edited by Alan J. Izenman, 237–280. Springer Texts in Statistics. New York, NY: Springer. https://doi.org/10.1007/978-0-387-78189-1_8.

Jaeger, Jeri J., and Robert D. Van Valin. 1982. "Initial Consonant Clusters in Yateé Zapotec." *International Journal of American Linguistics* 48 (2): 125–138.

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. Accessed August 24, 2022. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CMFD&dbname=CMFD201401&filename=1013324746.nh&uniplatform=NZKPT&v=k2I_D7XYr0iT6Ytn6U9S2vFmsOwljpv4-K1l2-8jkFwfQ0yGUw7CioYI52_FutHq.

Jia, Yuan, and Aijun Li. 2012. "Phonetic realization of accent from Chinese English learners in various dialectal regions." In *2012 8th International Symposium on Chinese Spoken Language Processing,* 296–300. Kowloon Tong, China: IEEE, December. ISBN: 978-1-4673-2507-3 978-1-4673-2506-6 978-1-4673-2505-9, accessed November 9, 2021. https://doi.org/10.1109/ISCSLP.2012.6423547. http://ieeexplore.ieee.org/document/6423547/.

Jongman, Allard. 2004. "Phonological and phonetic representations: The case of neutralization." In *Proceedings of the 2003 Texas Linguistics Society Conference: Coarticulation in Speech Production and Perception,* edited by Augustine Agwuele, Willis Warren, and Sang-Hoon Park. Somerville, MA: Cascadilla Proceedings Project.

Keating, Patricia, Marc Garellek, and Jody Kreiman. 2015. “Acoustic properties of different kinds of creaky voice.” In *Proceedings of the 18th International Congress of Phonetic Sciences.* Glasgow, UK: the University of Glasgow.

Kim, Hyunsoon, and Allard Jongman. 1996. “Acoustic and perceptual evidence for complete neutralization of manner of articulation in Korean.” *Journal of Phonetics* 24, no. 3 (July): 295–312. https://doi.org/10.1006/jpho.1996.0016.

Kirby, James. 2010. “Dialect experience in Vietnamese tone perception.” *The Journal of the Acoustical Society of America* 127 (6): 3749–3757. https://doi.org/10.1121/1.3327793.

———. 2011. “Vietnamese (Hanoi Vietnamese).” *Journal of the International Phonetic Association* 41:381–392. https://doi.org/10.1017/S0025100311000181.

Klatt, Dennis H., and Laura C. Klatt. 1990. “Analysis, synthesis, and perception of voice quality variations among female and male talkers.” *The Journal of the Acoustical Society of America* 87 (2): 820–857. https://doi.org/10.1121/1.398894.

Kuang, Jianjing. 2013. “Phonation in Tonal Contrasts.” PhD diss., University of California, Los Angeles.

———. 2017. “Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice.” *Journal of the Acoustical Society of America* 142:1693–1706.

———. 2018. “The influence of tonal categories and prosodic boundaries on the creakiness in Mandarin.” *The Journal of the Acoustical Society of America* 143 (6): EL509–EL515. https://doi.org/10.1121/1.5043094.

Kuang, Jianjing, and Patricia Keating. 2013. “Glottal articulations in tense vs lax phonation contrasts.” *The Journal of the Acoustical Society of America* 134, no. 5 (November): 4069–4069. https://doi.org/10.1121/1.4830854.

Kuo, Chen-Hsiu. 2013. “Perception and acoustic correlates of the Taiwanese tone sandhi group.” PhD diss., University of California, Los Angeles.

Ladefoged, Peter, and D. E. Broadbent. 1957. “Information conveyed by vowels.” *Journal of the Acoustical Society of America* 29:98–104. https://doi.org/10.1121/1.1908694.

Lai, Wenpan 赖文盼. 2016. “Jiyu Xiamen hua zuo Minnan fangyan shengdiao yanjiuj 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,* no. 4, 129–130. Accessed August 24, 2022. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2016&filename=YSKK201604118&uniplatform=NZKPT&v=SzxlJrhYH9kwNEydmNu85tiU6iQChwxJ03GkVDzAIs6e4waB_xcHYsgJX_r3jLhB.

Lalhminghlui, Wendy, and Priyankoo Sarmah. 2018. "Production and perception of rising tone sandhi in Mizo." In *6th International Symposium on Tonal Aspects of Languages (TAL 2018),* 114–118. ISCA. https://doi.org/10.21437/TAL.2018-23.

Leander, Anita J. 2008. "Acoustic correlates of fortis/lenis in San Francisco Ozolotepec Zapotec." Master's thesis, University of North Dakota.

Lee, Wai-Sum, and Eric Zee. 2009. "Hakka Chinese." *Journal of the International Phonetic Association* 39 (1): 107–111. https://doi.org/10.1017/S0025100308003599.

Li, Bing 李冰, 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 of study of citation tones and tone sandhi in Changsha Chinese]." *Journal of Hunan University (Social Sciences)* 20 (4): 107–112. Accessed August 24, 2022. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2006&filename=HDXB200604019&uniplatform=NZKPT&v=2y1mAjuFLfJAk2_o-UXRuG9ncebxJ-LIikFXmrGb9H4dKxve88sV_3p3oylK8dYe.

Li, Dongxiang 李冬香. 2005. "Hunan Ganyu yuyin yanjiu 湖南赣语语音研究 [A phonological study of Gan Dialect in Hunan Province]." PhD diss., Jinan University. Accessed August 24, 2022. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CDFD&dbname=CDFD9908&filename=2005142029.nh&uniplatform=NZKPT&v=q7Nj7CBBrBBXrm6ZgZXfCdIQFfU7Xjz1Fgt86t9JEfgHlQLV9OWIx8iboopBD1ws.

Li, Lilin 李立林. 2010. "Dongguan yueyu yuyin yanjiu 东莞粤语语音研究 [A phonological study of Yue dialects in Dongguan]." PhD diss., Jinan University. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CDFD&dbname=CDFD0911&filename=2010123719.nh&uniplatform=NZKPT&v=tZf6bKV19yYWewqy6KiDtrjdIeKqXjljZWUfQ7YQIwQoDxDQu39dfe1hcg60oosb.

Li, Rulong 李如龙, Shuangqing Zhang 张双庆, Bo Wan 万波, Yi Shao 邵宜, and Chunzhao Lian 练春招. 1992. *Kegan fangyan diaocha baogao [A Report on a Survey of the Kejia and Gan Dialects]* 客赣方言调查报告. Edited by Rulong Li 李如龙 and Shuangqing Zhang 张双庆. Xiamen: Xiamen University Press.

Li, Xiaoyu 李晓钰. 2007. "Xupu Xian Longtang fangyan yuyin yanjiu 溆浦县龙潭方言语音研究 [The phonological study on Longtan dialect in Xupu county]." Master's thesis, Hunan Normal University. Accessed August 24, 2022. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CMFD&dbname=CMFD2008&filename=2008133649.nh&uniplatform=NZKPT&v=XKyKBckxVmJro7W6w7lpQoUMAMugL9ZAUu62fxMgrqfGisKoxWHwoJx7KQyyxZIo.

Li, Yang. 2015. "Tone sandhi and tonal coarticulation in Fuzhou Min." In *Proceedings of the 18th International Congress of Phonetic Sciences.* Glasgow, U.K.: the University of Glasgow.

Lin, Xiaoxiao 林晓晓. 2011. “Wuyu Luqiao fangyan yuyin yanjiu 吴语路桥方言语音研究 [Research on the phonetics of the Luqiao Wu dialect].” Master’s thesis, Fujian Normal University. Accessed August 24, 2022. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CMFD&dbname=CMFD2012&filename=1012266916.nh&uniplatform=NZKPT&v=abHognL3m2npjeBeId9GYtc8U6K629prA4GA7PuB0pg2KUkuPTvlNOFT7ImTvQXC.

Liu, Jian 刘健. 2012. “Neimenggu zhongxibu jinyu yuyin yanjiu 内蒙古中西部晋语语音调查 [Phonetic investigation of Mid-Eastern Inner Mongolia].” Master’s thesis, Shanghai Normal University. Accessed August 24, 2022. http://cdmd.cnki.com.cn/Article/CDMD-10270-1012454764.htm.

Liu, Lili 刘俐李. 2008. “Jipin guiyi he diaoxi guizheng de fangyan shiyan 基频归一和调系规整的方言实验 [F0 normalization and tone adjustment in dialect experiments].” *Chinese Journal of Phonetics,* no. 1, 221–227.

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. Accessed August 24, 2022. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CMFD&dbname=CMFD201402&filename=1013245249.nh&uniplatform=NZKPT&v=5HIyN8FSwIE3J8WTNeVYPZZ6IIPpXy1SfH1OvY8GeADP9TrbCL5u2Ga5Pi%25mmd2FiNpmS.

Lyman, Larry, and Rosemary Lyman. 1977. “Choapan Zapotec phonology.” In *Studies in Otomanguean phonology,* edited by William R. Merrifield, 137–161. SIL International Publications in Linguistics 54. Mexico: Summer Institute of Linguistics / the University of Texas at Arlington. Accessed August 24, 2022. https://www.sil.org/resources/archives/8836.

Lyu, Xiang. 2019. “The phonological word in the Ningbo dialect.” *International Journal of Chinese Linguistics* 6 (1): 119–161. https://doi.org/10.1075/ijchl.19004.lyu.

MacLaury, Robert Ethan. 1970. “Ayoquesco Zapotec: Ethnography, phonology, and lexicon.” Master’s thesis, University of the Americas. Accessed August 24, 2022. https://colcat.calit2.uci.edu/docs/bio/MacLauryMastersThesis.pdf.

MacMillan, Neil A. 2002. “Signal Detection Theory.” In *Stevens’ Handbook of Experimental Psychology,* edited by Hal Pashler. Hoboken, NJ: John Wiley & Sons, Inc. https://doi.org/10.1002/0471214426.pas0402.

Maddieson, Ian. 1985. “Phonetic cues to syllabification.” In *Phonetic linguistics: essays in honor of Peter Ladefoged,* edited by V. Fromkin, 203–221. Orlando: Academic Press.

Maddieson, Ian, and Peter Ladefoged. 1985. “ “Tense” and “lax” in four minority languages of China.” *Journal of Phonetics* 13 (4): 433–454. https://doi.org/10.1016/S0095-4470(19)30788-0.

Marlett, Stephen A., and Velma B. Pickett. 1987. “The Syllable Structure and Aspect Morphology of Isthmus Zapotec.” *International Journal of American Linguistics* 53 (4): 398–422.

Matsui, Mayuki. 2011. “The Identifiability and Discriminability between Incompletely Neutralized Sounds: Evidence from Russian.” In *Proceedings of the 17th International Congress of Phonetic Sciences,* 1342–1345. Hong Kong. Accessed June 18, 2022. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/RegularSession/Matsui/Matsui.pdf.

Mei, Tsu-lin. 1970. “Tones and prosody in Middle Chinese and the origin of the rising tone.” *Harvard Journal of Asiatic Studies* 30:86–110. Accessed May 24, 2022. https://www.jstor.org/stable/2718766.

Merrill, Elizabeth D. 2008. “Tilquiapan Zapotec.” *Journal of the International Phonetic Association* 38 (01). https://doi.org/10.1017/S0025100308003344.

Michaud, A. 2004. “Final Consonants and Glottalization: New Perspectives from Hanoi Vietnamese.” *Phonetica* 61 (2-3): 119–146. https://doi.org/10.1159/000082560.

Nellis, Donald G., and Barbara E. Hollenbach. 1980. “Fortis versus Lenis in Cajonos Zapotec Phonology.” *International Journal of American Linguistics* 46 (2): 92–105. https://doi.org/10.1086/465639.

Nelson, Julia Louise. 2004. “Tone and glottalization in San Juan Mixtepec Zapotec.” Master’s thesis, University of Texas at Arlington. Accessed August 24, 2022. https://www.sil.org/resources/archives/9450.

Norman, Jerry. 1969. “The Kienyang Dialect of Fukien.” ISBN: 9781084827431. PhD diss., University of California, Berkeley.

———. 1988a. *Chinese.* Cambridge: Cambridge University Press.

———. 1988b. “The She Dialect of Luoyuan County.” *Bulletin of the Institute of History and Philology* 59:353–367.

Oakden, Christopher. 2017. “Checked Tone Merger in the Nanjing Dialect: An Acoustic Analysis.” In *Proceedings of the 29th North American Conference on Chinese Linguistics,* edited by Lan Zhang, 1:141–152. Columbus, OH.

Oliva-Juarez, Gabriela, Fabiola Martinez-Licona, Alma Martinez-Licona, and John Goddard-Close. 2014. “Identification of Vowel Sounds of the Choapan Variant of Zapotec Language.” In *Nature-Inspired Computation and Machine Learning,* edited by Alexander Gelbukh, Félix Castro Espinoza, and Sofía N. Galicia-Haro, 252–262. Cham: Springer International Publishing. ISBN: 978-3-319-13650-9.

Pan, Ho-hsien. 2017. "Glottalization of Taiwan Min checked tones." *Journal of the International Phonetic Association* 47 (1): 37–63. https://doi.org/10.1017/S0025100316000281.

Pan, Ho-hsien, Hsiao-tung Huang, and Shao-ren Lyu. 2016. "Coda stop and Taiwan Min checked Tone sound changes." In *Interspeech 2016,* 1011–1015. ISCA. https://doi.org/10.21437/Interspeech.2016-597.

Pan, Ho-hsien, and Shao-ren Lyu. 2021. "Taiwan Min Nan (Taiwanese) Checked Tones Sound Change." In *Interspeech 2021,* 2641–2645. ISCA. https://doi.org/10.21437/Interspeech.2021-672.

Parker, Steve. 2001. "Non-optimal onsets in Chamicuro: an inventory maximised in coda position." *Phonology* 18 (3): 361–386. https://doi.org/10.1017/S0952675701004122.

Peirce, Jonathan, Jeremy R. Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman, and Jonas Kristoffer Lindeløv. 2019. "PsychoPy2: Experiments in behavior made easy." *Behavior Research Methods* 51 (1): 195–203. https://doi.org/10.3758/s13428-018-01193-y.

Peng, Shu-hui. 1996. "Phonetic implementation and perception of place coarticulation and tone sandhi." PhD diss., The Ohio State University.

———. 2000. "Lexical versus 'phonological' representations of Mandarin Sandhi tones." In *Acquisition and the Lexicon,* edited by Michael M. Broe and Janet B. Pierrehumbert, 152–167. Papers in laboratory phonology 5. Cambridge: Cambridge University Press.

Pickett, Velma B., María Villalobos Villalobos, and Stephen A. Marlett. 2010. "Isthmus (Juchitán) Zapotec." *Journal of the International Phonetic Association* 40 (3): 365–372. https://doi.org/10.1017/S0025100310000174.

Pinson, Thomas, and Jacqueline Pinson. 2000. *An introduction to Kunming Hua.* 3rd. Dallas, Texas: SIL International East Asia. Accessed January 15, 2022. https://halshs.archives-ouvertes.fr/halshs-01383557.

Prajwala, T.R. 2015. "A Comparative Study on Decision Tree and Random Forest Using R Tool." *International Journal of Advanced Research in Computer and Communication Engineering,* 196–199. https://doi.org/10.17148/IJARCCE.2015.4142.

Qian, Nairong 钱乃荣. 1990. "Ningbo fangyan xinpai yinxi fenxi 宁波方言新派音系分析 [Analysis of the phonology of Ningbo dialect by younger generation]." *Studies in Language and Linguistics* 语言研究 1. Accessed August 24, 2022. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD9093&filename=YYYJ199001011&uniplatform=NZKPT&v=p3x9g2u0-cpe_O18AOfuvg2maH3gvz6GJa3s6HnAFJQ1gndNwc1YHhHcWXilE40C.

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* 34 (January). ISSN: 1043-3805, accessed November 2, 2021. https://doi.org/10.17161/KWPL.1808.12864. http://hdl.handle.net/1808/12864.

Ramsey, Jim, and Brian Ripley. 2017. *Pspline: Penalized smoothing splines,* June. Accessed November 17, 2021. https://CRAN.R-project.org/package=pspline.

Ratliff, Martha. 2010. *Meaningful tone: a study of tonal morphology in compounds, form classes, and expressive phrases in white Hmong.* OCLC: ocn497574253. DeKalb: Northern Illinois University Press. ISBN: 978-0-87580-636-5.

———. 2015. *Tonoexodus, Tonogenesis, and Tone Change.* Edited by Patrick Honeybone and Joseph Salmons. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199232819.013.021.

Rietveld, A.C.M., and Ulrich Hans Frauenfelder. 1987. “The Effect of Syllable Structure on Vowel Duration.” In *Proceedings of the 11th International Congress of Phonetic Sciences,* 28–31. Tallinn. Accessed August 24, 2022. https://archive-ouverte.unige.ch/unige:85564.

Riggs, Sylvia. 2011. “Beginning Tone Analysis for Amatlán Zapotec.” *SIL-Mexico Electronic Working Papers,* SIL-Mexico Electronic Working Papers, no. 10, 15. Accessed August 24, 2022. https://www.sil.org/resources/archives/51459.

Roettger, Timo B., Bodo Winter, Sven Grawunder, James Kirby, and Martine Grice. 2014. “Assessing incomplete neutralization of final devoicing in German.” *Journal of Phonetics* 43:11–25. https://doi.org/10.1016/j.wocn.2014.01.002.

Sagart, Laurent. 1986. “On the departing tone.” *Journal of Chinese Linguistics* 14 (1): 90–113. Accessed May 24, 2022. https://www.jstor.org/stable/23754220.

———. 1990. “Un dialecte du nord Jiangxi: Wuning.” *Cahiers de linguistique - Asie orientale* 19 (2): 143–193. https://doi.org/10.3406/clao.1990.1328.

———. 1993. “Austronesian Final Consonants and the Origin of Chinese Tones.” *Oceanic Linguistics Special Publications,* no. 24, 47–59. Accessed May 24, 2022. http://www.jstor.org/stable/20006747.

Santen, J.P.H. van. 1992. “Contextual effects on vowel duration.” *Speech Communication* 11 (6): 513–546. https://doi.org/10.1016/0167-6393(92)90027-5.

Seyfarth, Scott, and Marc Garellek. 2018. “Plosive voicing acoustics and voice quality in Yerevan Armenian.” *Journal of Phonetics* 71 (November): 425–450. https://doi.org/10.1016/j.wocn.2018.09.001.

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. Accessed August 24, 2022. https://cdmd.cnki.com.cn/Article/CDMD-10319-1013105547.htm.

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,* no. 4, 296–304. Accessed August 24, 2022. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2007&filename=FYZA200704005&uniplatform=NZKPT&v=wf-ZZcpwLvonVi4nagZA0mZiAFcKj3v8oIFGbbKdwDY7714Rn9aAKmJL9UoM3kYH.

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,* no. 1, 1–14. Accessed August 24, 2022. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CCJD&dbname=CCJDLAST2&filename=NKYK201001003&uniplatform=NZKPT&v=p0Tr3X2zKSpMI5HT22HT-fCWqMvnnNtHfVGTPv5Y7Ai_Ty5Ja3V8qyIBVW9LJ02K.

Shi, Jun 施俊. 2012. "Zhejiang Yiwu fangyan rusheng shuhua tanxi 浙江义乌方言入声舒化探析 [On the softening in entering tone of Yiwu Dialects, Zhejiang province]." *Dialect* 方言 1:83–90. Accessed August 24, 2022. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2012&filename=FYZA201201014&uniplatform=NZKPT&v=qRY3_c0V8uVCv3c5O-18YEvC4yQZ24KPtdDMwIXLLWUr7bzzDAKo50wbjVrmk3Dp.

Shih, Stephanie S., and Sharon Inkelas. 2019. "Autosegmental Aims in Surface-Optimizing Phonology." *Linguistic Inquiry* 50:137–196. https://doi.org/10.1162/ling_a_00304.

Shue, Yen-Liang, Patricia Keating, Chad Vicenik, and Kristine M. Yu. 2011. "VoiceSauce: A program for voice analysis." In *Proceedings of the 17th International Congress of Phonetic Science,* 1846–1849. Hong Kong, August.

Silverman, B. W. 1986. *Density estimation for statistics and data analysis.* Monographs on statistics and applied probability 26. London: Chapman & Hall.

Slowiaczek, Louisa M., and Helena J. Szymanska. 1989. "Perception of word-final devoicing in Polish." *Journal of Phonetics* 17 (3): 205–212. https://doi.org/10.1016/S0095-4470(19)30430-9.

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,* no. 2, 166–172. Accessed August 24, 2022. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2009&filename=NJSD200902038&uniplatform=NZKPT&v=l5kAM0Iw8dEgT8wCtY5IrXI-R3W-uwiqj8pgFT0OpGurM5sp6yW-YxmRmJgMRbXu.

Sonnenschein, Aaron Huey. 2004. “A descriptive grammar of San Bartolomé Zoogocho Zapotec.” PhD diss., University of Southern California.

Speck, Charles H. 1978a. “Texmelucan Zapotec suprasegmental phonology.” *Work Papers of the Summer Institute of Linguistics, University of North Dakota Session* 22 (1). https://doi.org/10.31356/silwp.vol22.09.

———. 1978b. “The Phonology of Texmelucan Zapotec Verb Irregularity.” Master’s thesis, University of North Dakota. Accessed August 24, 2022. https://commons.und.edu/theses/2660.

———. 1984. “The Phonology of the Texmelucan Zapotec Verb.” Publisher: University of Chicago Press, *International Journal of American Linguistics* 50 (2): 139–164. Accessed December 26, 2021. https://www.jstor.org/stable/1265602.

Stanislaw, Harold, and Natasha Todorov. 1999. “Calculation of signal detection theory measures.” *Behavior Research Methods, Instruments, & Computers* 31:137–149.

Strobl, Carolin, Anne-Laure Boulesteix, Achim Zeileis, and Torsten Hothorn. 2007. “Bias in random forest variable importance measures: Illustrations, sources and a solution.” *BMC Bioinformatics* 8 (1): 25. https://doi.org/10.1186/1471-2105-8-25.

Stubblefield, Morris, and Barbara E. Hollenbach. 1991. *Gramática zapoteca: Zapoteca de Mitla, Oaxaca.* Translated by Sara Arjona de Watson. Instituto Lingüístico de Verano, A.C. Accessed August 24, 2022. https://www.sil.org/resources/archives/88999.

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,* no. 4, 19–23. Accessed August 24, 2022. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLASN2019&filename=YWLY201604007&uniplatform=NZKPT&v=VSM7oxLJkqm9-zTTT9doAtYGQQ5ww3qxS-gq0D-5Q9Vda52r5ASLVK59K8B-B8Up.

Sun, Yuqing 孙玉卿, and Maolin Wang 王茂林. 2006. “Datong fangyan “shusheng cuhua” yu qingsheng yinjie de guanxi 大同方言 “舒声促化” 与轻声音节的关系 [An analysis of checkedization and neutral tone in Tatong dialect].” *Journal of Shanxi University (Philosophy & Social Science)* 29 (3): 112–115. Accessed August 24, 2022. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2006&filename=SXDD200603025&uniplatform=NZKPT&v=423BPmMO2EA0s1MlVzmphBGUk_DQaM_6zXNmdQ2sD2gvp19l_dBK6W5nrfP_OQ3X.

Sung, Margaret M. Y. 1973. “A study of literary and colloquial Amoy Chinese.” *Journal of Chinese Linguistics* 1 (3): 414–436. Accessed January 15, 2022. https://www.jstor.org/stable/23752861.

Tan, Jieying 谭洁莹. 2016. "Yue fangyan Taishan hua yuyin tezheng de shiyan yuyinxue yanjiu 粤方言台山话语音特征的实验语音学研究 [An acoustic study on the phonetic features of Taishan dialect, Cantonese]." Master's thesis, Jinan University. Accessed August 24, 2022. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CMFD&dbname=CMFD201701&filename=1016733648.nh&uniplatform=NZKPT&v=_TUf0Xbaqg4CGMQbr83YTjPBTosMEFB5ME9JS7SY9EPawlltCew6QbznZv-V4Ufp.

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. Accessed August 24, 2022. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CMFD&dbname=CMFD201501&filename=1014347991.nh&uniplatform=NZKPT&v=iPwMUEH7qe0ikdrD9kZAlc%25mmd2F2xJSf61Ck42RSLQm0obCKMmGjvuBULmEjVjJJHQb1.

Teodocio Olivares, Amador. 2009. "Betaza Zapotec Phonology: Segmental and Suprasegmental Features." Master's thesis, University of Texas at Austin. Accessed August 24, 2022. https://repositories.lib.utexas.edu/handle/2152/19162.

Thurgood, Graham. 1976. "The origins of Burmese creaky tone." PhD diss., University of California, Berkeley.

Ting, Pang-hsin. 1980. "The Tan-chou dialect of Hainan." *Cahiers de Linguistique - Asie Orientale* 8 (1): 5–27. Accessed August 24, 2022. https://www.persee.fr/doc/clao_0153-3320_1980_num_8_1_1079.

Tung, T'ung-ho 董同龢, Yung-lang Chao 趙榮琅, and Ya-shiu Lan 藍亞秀. 1967. *Ji Taiwan de yizhong Minnan hua* 記臺灣的一種閩南話 *[A south Min dialect of Taiwan].* A 24. Taipei, Taiwan: Institute of History / Philosophy, Academic Sinica.

Uchihara, Hiroto, and Ambrocio Gutiérrez. 2019. "El texto Don Crescencio: ilustración del sistema tonal del zapoteco de Teotitlán del Valle." *Tlalocan* 24:127–155. https://doi.org/10.19130/iifl.tlalocan.2019.487.

———. 2020. "Subject and agentivity in Teotitlán Zapotec." *Studies in Language* 44 (3): 548–605. https://doi.org/10.1075/sl.18025.uch.

Venables, W. N., Brian D. Ripley, and W. N. Venables. 2002. *Modern applied statistics with S.* 4th ed. Statistics and computing. New York: Springer. ISBN: 978-0-387-95457-8.

Warner, Natasha, Allard Jongman, Joan Sereno, and Rachèl Kemps. 2004. "Incomplete neutralization and other sub-phonemic durational differences in production and perception: evidence from Dutch." *Journal of Phonetics* 32 (2): 251–276. https://doi.org/10.1016/S0095-4470(03)00032-9.

Watkins, Justin. 2000. "Notes on creaky and killed tone in Burmese." *SOAS Working papers in Linguisitics and Phonetics* 10:139–149. Accessed June 28, 2022. https://eprints.soas.ac.uk/1948/1/SOASWorkingPaper2000%20-%20Creaky%20and%20killed%20tone%20in%20Burmese.pdf.

———. 2001. "Burmese." *Journal of the International Phonetic Association* 31 (2): 291–295. https://doi.org/10.1017/S0025100301002122.

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. Accessed August 24, 2022. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CMFD&dbname=CMFD201601&filename=1015720187.nh&uniplatform=NZKPT&v=taJQvNmlO3IpPKCgVcRTtW2y8H1jQJifrIuyGsKi9UELK4n6ZFJ2i5dENKvLXYqr.

Woon, Wee-Lee. 1979a. "A synchronic phonology of Hainan dialect: Part I." Publisher: [Chinese University Press, Project on Linguistic Analysis], *Journal of Chinese Linguistics* 7 (1): 65–100. Accessed January 15, 2022. https://www.jstor.org/stable/23753034.

———. 1979b. "A synchronic phonology of Hainan dialect: Part II." Publisher: [Chinese University Press, Project on Linguistic Analysis], *Journal of Chinese Linguistics* 7 (2): 268–302. Accessed January 15, 2022. https://www.jstor.org/stable/23752923.

Wu, Bo. 2018. "An acoustic analysis of vowels in checked syllables in Chinese." *Chinese Journal of Acoustics,* no. 04, 491–502. https://doi.org/10.15949/j.cnki.0217-9776.2018.04.009.

Xia, Liping, and Fang Hu. 2016. "Vowels and diphthongs in the Taiyuan Jin Chinese dialect." In *Interspeech 2016,* 993–997. ISCA. https://doi.org/10.21437/Interspeech.2016-249.

Xiapu County Bureau of Statistics. 2021. *Xiapu xian diqici quanguo renkou pucha gongbao [Xiapu County seventh national census report].* Accessed August 24, 2022. http://www.xiapu.gov.cn/zwgk/zfxxgkzdgz/tjxx/tjgb/202106/t20210628_1491098.htm.

Xing, Xiangdong 邢向东, and Wanchun Meng 孟万春. 2006. "Shaanbei Ganquan Yanchang fangyan rushengzi duyin yanjiu 陕北甘泉、延长方言人声字读音研究 [On the pronunciation of entering tone characters in Ganquan and Yanchang dialects in northern Shaanxi]." *Chinese Language,* no. 5, 444–451. Accessed August 24, 2022. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2006&filename=YWZG200605011&uniplatform=NZKPT&v=J4AHOlbFZukas_X-6IooOyGPOkhi-_pqKiu4dkxejIq0YlNxBXeitlln934OAJdv.

Xu, Chenchen. 2020. "The phonology and phonetics of Rugao syllable contraction." PhD diss., Michigan State University. https://doi.org/10.25335/STVG-3783.

Yang, Shifeng 楊時逢. 1987. *Lizhuang fangyan ji* 李莊方言記 *[The dialects of Lizhuang].* Taipei: Institute of History / Philosophy, Academic Sinica. Accessed August 24, 2022. http://ir.sinica.edu.tw/handle/201000000A/6213.

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,* no. 1, 130–133. Accessed August 24, 2022. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD7984&filename=YWYJ198201021&uniplatform=NZKPT&v=GCPiTBeyAbjxyVRY7Uyiud0lSxsB_lSQ_gwQIjkr2bka1x4-aQbvam_UqGiTq8pj.

Yang, Yongkai, and Ying Chen. 2018. "Effects of Entering Tone on Vowel Duration and Formants in Nanjing Dialect." In *Studies on Speech Production,* edited by Qiang Fang, Jianwu Dang, Pascal Perrier, Jianguo Wei, Longbiao Wang, and Nan Yan, 10733:146–157. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-00126-1_14.

Yu, Kristine M., and Hiu Wai Lam. 2014. "The role of creaky voice in Cantonese tonal perception." *The Journal of the Acoustical Society of America* 136 (3): 1320–1333. https://doi.org/10.1121/1.4887462.

Yuan, Jiahong, and Yiya Chen. 2014. "3rd tone sandhi in standard Chinese: A corpus approach [大语料库中的普通话上声连读变调]." *Journal of Chinese Linguistics* 42 (1): 218–236. Accessed June 18, 2022. https://www.jstor.org/stable/23753947.

Zee, Eric, and Liejiong Xu. 2017. "Shanghainese." In *The Sino-Tibetan Languages,* Second edition, edited by Graham Thurgood and Randy J. LaPolla, 185–200. Routledge language family series. London ; New York: Routledge. ISBN: 978-1-138-78332-4.

Zhang, Guangming 张光明. 2006. "Xinzhou fangyan de shusheng cuhua xianxiang 忻州方言的舒声促化现象 [Non-Ru tone becoming Ru tone in Xinzhou dialect]." *Yuwen Yanjiu* 语文研究 2:54–57. Accessed August 24, 2022. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2006&filename=YWYJ200602012&uniplatform=NZKPT&v=rTI0Y8c3LsNK5JGzaC8O8AdoKyD0swrCTM9tFzkdm0RZmREb06eJwwJEft76RuLm.

Zhang, Jingfen. 2017. "Tonal patterns and tone evolution of Chaoshan Chinese." PhD diss., Hong Kong University of Science and Technology.

Zhang, Jisheng. 2006. *The phonology of Shaoxing Chinese.* Utrecht: LOT. ISBN: 978-90-76864-90-7.

Zhang, Shengkai. 2014. “Chinese Pingjiang dialect.” In *Grammatical Sketches from the Field,* edited by Toshihide Nakayama, Noboru Yoshioka, and Kosei Otsuka, 67–107. Tokyo: Research Institute for Languages / Cultures of Asia / Africa (ILCAA), Tokyo University of Foreign Studies. http://repository.tufs.ac.jp/bitstream/10108/88522/1/GrammaticalSketch_Final.pdf.

Zhang, Zhaoyan. 2016. “Cause-effect relationship between vocal fold physiology and voice production in a three-dimensional phonation model.” *The Journal of the Acoustical Society of America* 139 (4): 1493–1507. https://doi.org/10.1121/1.4944754.

Zhengzhang, Shangfang 郑张尚芳. 2003. *Shanggu Yinxi* 上古音系 *[Old Chinese Phonology].* 1st ed. Shanghai: Shanghai Educational Publishing House.

Zhu, Min. 2015. “Studies on Chinese Dialect Grammar: The Gan Dialect of Chongyang.” PhD diss., Eötvös Loránd University.

Zhu, Xiaonong 朱晓农. 2004. “Jipin guiyihua - ruhe chuli shengdiao de suiji chayi? 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 (2): 3–19. http://journal15.magtechjournal.com/Jwk_yykx/CN/Y2004/V3/I2/3.

Zhu, Xiaonong 朱晓农, and Lei Jiao 焦磊. 2011. “Duandiao wu se wei – Baogao yili teshu de rusheng leixing 短调无塞尾 – 报告一例特殊的入声类型 [Short tones without a plosive coda: A new type of the entering tone syllable].” *Yuyan Yanjiu Jikan* 语言研究集刊 1:142–147. Accessed August 24, 2022. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CCJD&dbname=CCJDLAST2&filename=YJJK201100010&uniplatform=NZKPT&v=DRbBuFuCdWBTw2oURy1tD7RvPHe0cvP2DXoB9Y5UpSR_2sPtn7eXcg3_eqNbPYf4.

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* 中国语文*,* no. 4, 324–338. Accessed August 24, 2022. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2008&filename=YWZG200804007&uniplatform=NZKPT&v=Bu-eB5Omdi4kcpQO6y-Sw6wtkb06DDs_OhU-7scdVxkwrg8TYeg9JgiODsyRRBwM.

Zurita Sánchez, Emiliano, Michael Ward, and Stephen A. Marlett. 2008. “Zapoteco de Santa Catarina Quioquitani.” In *Ilustraciones fonéticas de lenguas Amerindias,* edited by Stephen A. Marlett, 9. Lenguas de las Américas.