Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

Perceptual Cues to Checked Tones in Shanghai Chinese

Chengjia Ye[1], *Paul Boersma*[2]

[1]Donders Institute for Brain, Cognition, and Behaviour, Radboud University, the Netherlands
[2]University of Amsterdam, the Netherlands
chengjia.ye@donders.ru.nl, paul.boersma@uva.nl

Abstract

Shanghai Chinese has a complicated tone system conditioned by phonological onset voicing and syllable structure. Checked tones in Shanghai Chinese are acoustically realized with tenser phonation, shorter duration and a more central vowel space than their unchecked counterparts. The present study examines whether these cues also play a role in speech perception. A three-alternative forced-choice experiment was conducted in which 40 native listeners chose among three Chinese characters representing an open /CV/, a nasal-ending /CVN/ and a glottal-ending /CVʔ/ syllable according to the word they heard. Results show main effects of phonation type, vowel quality and duration as well as the interaction between the former two, on both decisions and reaction time. For reaction time, two-way interactions between duration and the other two cues were also detected. From the non-overlapping confidence intervals of the three main effects we tentatively conclude that tenser phonation is the primary perceptual cue to checked tones, while a more central vowel space and shorter duration are secondary, in line with findings in other tonal languages like Vietnamese and Cantonese. Additionally, we found that the low vowel pair /ɐ/ and /ɑ/ and high register tones tend to bias perception towards checked tones.

Index Terms: speech perception, tone, phonation, vowel duration, Shanghai Chinese

1. Introduction

Fundamental frequency (*f*0) is usually a primary perceptual cue to lexical tones, as is the case in Mandarin Chinese. However, in many tonal languages, *f*0 is often accompanied by other cues, including but not limited to phonation type, onset voicing and duration, rhyme structure and duration as well as vowel quality.

Phonation types (or voice quality, see [1], a review on phonation contrasts), for instance, are an essential cue to tone identification, particularly in languages in which the *f*0 contour patterns of several tones closely resemble one another. Three among the seven lexical tones in Green Hmong share similar duration, *f*0 height and a falling contour, while intended breathy and creaky tones are identified with higher accuracy than the modal tone [2]. In North Vietnamese, which has six lexical tones but only four *f*0 contours, voice quality (modal vs. mid-laryngealized) plays a key role in tone identification; final glottalization even overrides *f*0 in the perception of a low tone [3]. Similarly, creakiness is also the primary cue to a low tone in Cantonese by increasing its accuracy in identification and biasing another tone with a similar *f*0 contour towards it [4].

Duration is another common cue but usually co-occurs with others. Tones in Vietnamese with final glottalization are shorter than other tones [5]. However, duration is often considered secondary (e.g., [3], [6]) as a result of glottalization. In a related vein, three of the six tones in Cantonese have checked allotones closed with a stop coda and thus have shorter duration than their unchecked counterparts, whereas their *f*0 contours are nearly the same [7]. These checked allotones originated from the checked tones in Middle Chinese, whose remnants can also be seen in Shanghai Chinese, a subvariant of Wu.

As a tonal language, today's Shanghai Chinese has a rather complicated tone system with five tones [8], [9], in which T1, T2 and T4 reside in the high register, while T3 and T5 reside in the low one. The high tones are associated with phonemically voiceless onsets and start at a higher *f*0, whilst the low tones are associated with voiced onsets and start with a lower *f*0 [10]. However, voiced obstruents in the initial position of a prosodic unit have been acoustically devoiced in most cases as a result of diachronic sound change [9]−[11]. Instead, they are mainly realized with shorter closure or constriction duration than voiceless (unaspirated) ones [12] and with slight breathiness at the release as well as the following vowel onset [13], [14]. More recently, the phonation contrast between the two registers tends to fade away among the young generation [14], [15].

In the 19th century, each register comprised four pitch contours, yielding eight tones in accordance with Middle Chinese [16]. Nowadays, only five remain after diachronic tone mergers. Compatible with open /(C)V/ or nasal-ending /(C)VN/ syllables, T1 is a high-falling tone with no counterpart in the low register, T2 is high-rising and T3 is low-rising. By contrast, T4 and T5 only occur in syllables closed with a glottal stop /ʔ/ on a phonological level, and are known as *checked tones*. Table 1 summarizes the registers, pitch contours and syllable structures of these five tones.

Table 1: *The five tones in Shanghai Chinese sorted by their registers, pitch contours and syllable structures.*

syllable structure →		/(C)V(N)/		/(C)Vʔ/
register	onset	falling	rising	checked
High	[−voice]	T1	T2	T4
Low	[+voice]	–	T3	T5

However, the glottal coda is often acoustically realized not as a full but as an incomplete glottal closure, bringing tenser phonation (stronger vocal constriction and slightly higher *f*0) than unchecked tones to the nucleus, as demonstrated by both acoustic and articulatory measures of voice quality of T3 versus T5 [17]. In addition, syllables with a checked tone (*checked syllables*) are shorter than those with an unchecked tone (*unchecked syllables*). Chen's [18] production study compared vowels in unchecked open syllables with vowels in checked syllables when embedded in a sentence and revealed that the former are about 1.5 times longer and exhibit a more extensive vowel space than the latter, hence a difference in vowel quality.

10.21437/SpeechProsody.2024-79

The current study aims to investigate whether the three acoustic cues shown by [17], [18], i.e., phonation type, duration and vowel quality would also play a role in the perception of checked tones T4 and T5 in monosyllabic Shanghai Chinese words. Based on the *f*0-contour similarities between T2 and T4 and between T3 and T5 [9], [19], [20], we attempt to understand under what circumstances T2 and T3 are biased towards T4 and T5 respectively, and vice versa. For this purpose, we conducted a three-alternative forced-choice (3AFC) word identification task, in which native listeners heard a monosyllabic word and then had to choose among three Chinese characters representing the three possible syllable structures (open, nasal-ending and glottal-ending).

2. Methods

2.1. Participants

Forty native listeners of Shanghai Chinese (27 females and 13 males, average reported age = 27.00 years, *SD* = 9.53 years) with normal hearing and no reading difficulties participated in the task. All were born and had spent most of their lifetime in the urban (26 participants) or close-by suburban areas in Shanghai. They all self-reported to actively communicate in Shanghai Chinese in their daily lives. All participated in this task on a voluntary basis with consent obtained prior to the test.

2.2. Stimuli

The monosyllabic words were produced in isolation by a male native speaker with an urban dialect of Shanghai Chinese. Alveolar fricatives were used as onsets for all stimuli. Since voiceless onsets only co-occur with high tones, while voiced onsets only co-occur with low tones, voiceless /s/ was used for T2 and T4 (T1 is excluded from this study) and voiced /z/ was used for T3 and T5. Since the voiceless onset /s/ is longer than the voiced onset /z/ (as high tones have longer onsets compared with low tones [12]), we fixed the frication duration of /s/ and /z/ at 200 ms and 150 ms respectively, for all stimuli.

Two vowel pairs with relatively close F1 and F2 values were chosen to construct stimuli: the middle pair /o/–/ɔ/ and the low pair /ɐ/–/ɑ/. We refer to /o/ and /ɐ/ as *checked vowels*, as they are always followed by a glottal coda and hence bear the checked tones T4 and T5; they are somewhat centralized in terms of the acoustic vowel space. We refer to /ɔ/ and /ɑ/ as *unchecked vowels* as they only occur in open or nasal-ending syllables (with an unchecked tone T2 or T3 in this study); these vowels are more peripheral in the vowel space.

All four vowels were produced with both modal phonation, which ought to be associated with unchecked tones, and tense phonation (as the result of an incomplete final /ʔ/), which is a marker of checked tones. By applying the eight combinations to the two registers (High: higher *f*0 + voiceless onset vs. Low: lower *f*0 + voiced onset), we created 16 syllables. Among these, eight syllables had well-formed Shanghai Chinese phonetic implementations, as their vowel quality was congruent with phonation type: in [soʔ, sɐʔ] with T4 and [zoʔ, zɐʔ] with T5, both cues should indicate a checked tone; and in [sɔ, sɑ] with T2 and [zɔ, zɑ] with T3, both cues should indicate an unchecked tone. The other eight syllables were phonetically ill-formed, because the two acoustic cues they contain are contradictory: four checked vowels mismatched with modal phonation (*[so, sɐ] with T2 and *[zo, zɐ] with T3), and four unchecked vowels mismatched with tense phonation (*[sɔʔ, sɑʔ] with T4 and *[zɔʔ, zɑʔ] with T5).

Vowel duration in these 16 syllables was manipulated in Praat [21]. The shortest duration for the /o/–/ɔ/ pair was set as 130 ms, and that for /ɐ/–/ɑ/ as 120 ms. The longest duration was 410 ms for the former and 450 ms for the latter. These four values were approximately the average of these four vowels from a pilot monosyllabic word production study. In each of these two continua, three extra duration values were inserted, yielding five values on each continuum whose logarithms were equidistant, as shown in Figure 1. However, the largest value in each continuum was eliminated because the checked vowels /o/ and /ɐ/ extended for over 3 times sounded unnatural.

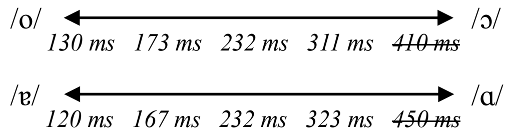

Figure 1: *Duration values on the two vowel continua.*

Altogether, there were 64 stimuli (2 tone registers × 4 vowels [i.e., two on each continuum] × 2 phonation types × 4 duration steps). Another 16 nasal-ending words (2 registers × 2 rhymes [commonly notated as /oŋ/ and /ɑ̃/] × 4 duration steps) served as fillers. Each stimulus and filler was played twice during the experiment. Each participant thus heard 160 words in total, whose order was randomized during the test.

2.3. Procedure

The 3AFC word identification task was conducted fully online. Participants were instructed to sit in a quiet room with their headphones on. During the experiment, they listened to the sound and then decided which Chinese character on the screen was (best) associated with the syllable they heard by pressing the button corresponding to the character as quickly as possible. Decisions and reaction times (RTs) were recorded.

At the beginning of each trial, there was a black fixation cross in the center of the white screen. The fixation cross disappeared after 500 ms, and the audio of the syllable was then played automatically. Then three Chinese characters appeared horizontally in the middle of the screen at the same time. The one on the left represented an unchecked open syllable, the one in the middle represented an unchecked nasal-closed syllable and the one on the right represented a checked glottal-ending syllable; for instance, the three options for stimulus [zo] with T3 and duration step 2 were /zɔ/ with T3, /zoŋ/ with T3, and /zoʔ/ with T5 (all legit). Their positions on the screen were fixed, and they corresponded to the "A", "S" and "D" keys on the keyboard respectively. Each syllable was played twice with two different sets of characters (homophones) during the entire experiment. Once the decision was made, the trial ended.

Prior to the test, there was a training session that consisted of five trials to familiarize the participants with the voice of the speaker and the experiment setting. Immediate feedback on their decision was given only during the training session but not during the test. Instructions were written in Chinese. The entire experiment lasted around 15 to 20 minutes per participant.

3. Results

The distribution of all 6400 responses (40 participants × 160 items) is summarized in Figure 2, categorized according to the presence of two auditory cues, tense phonation [±tense] and final nasality [±nasal], in the stimuli. Most stimuli acoustically realized with a nasal ending (and modal phonation) (96.95%) were identified as nasal-closed syllables regardless of duration. Functioning as fillers in the task, all stimuli with final nasality were excluded from further analysis. A large number of stimuli with tense phonation (87.30%) were identified as glottal-ending (i.e., checked) syllables regardless of other factors.

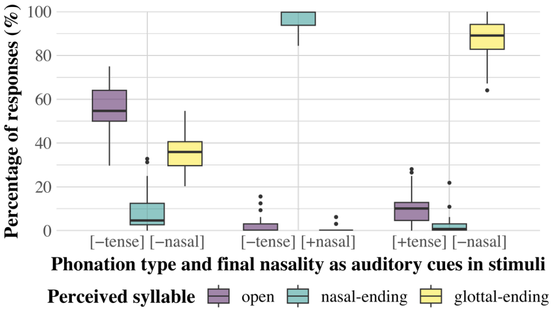

Figure 2: *Distribution of decisions per stimulus type.*

All stimuli perceived as nasal-closed syllables were also excluded from the analysis of decisions. Thus, there were 4833 (94.39%) responses remaining for the analysis of decisions. All responses with an RT longer than 5125 ms were excluded from the analysis of RT as outliers, leaving 4630 responses.

3.1. Decisions

A generalized linear mixed-effects regression (*glmer*) model was constructed in R version 4.3.1 [22] with the *lmerTest* package [23] to analyze the 4833 binary outcomes in which a stimulus is perceived as either an unchecked (0) or a checked tone (1) across various conditions.

Vowel pairs (either the /o/−/ɔ/ pair, coded as −0.5, or the /ɐ/−/ɑ/ pair, coded as +0.5) and tone registers (the high tones T2 and T4 were coded as −0.5, while the low tones T3 and T5 were coded as +0.5) were included as fixed effects. Since neither were involved in our research question, we did not consider their interaction. The three factors that function as acoustic cues to checked tones [17], [18], together with their interactions, were incorporated into the model. Phonation type (tense = +0.5; modal = −0.5) and vowel quality (checked vowels = +0.5; unchecked vowels = −0.5) were both binary factors. Duration was an ordinal variable with four levels, on which orthogonal polynomial coding was applied to explore the potential linear (the four levels were coded as −1/2, −1/6, +1/6, +1/2 in ascending order), quadratic (+1/2, −1/2, −1/2, +1/2) and cubic (−1/6, +1/2, −1/2, +1/6) trends. There was a random intercept per participant. More complicated models failed to converge; a series of likelihood ratio tests based on the *lrtest()* function in the *lmtest* package [24] found no significant effect of random slopes on the performance of the model.

Five main effects and a two-way interaction were revealed. The first main effect is vowel pair: Shanghai Chinese native listeners are estimated to have a 1.73 to 2.45 times greater odds (the 95% CI of the odds ratio, based on the exponentiated coefficients) of perceiving a stimulus as a checked tone (rather than an unchecked tone) when the vowel falls in the /ɐ/–/ɑ/ category than in the /o/–/ɔ/ category (point estimate of the odds ratio = 2.06, z = +8.221, p = 2.0 × 10^{-16}). Tone register also plays a role in the perception of checked tones: people have a 1.24 to 1.74 times greater odds of perceiving a syllable with a high register tone T2 or T4 together with the voiceless onset /s/ as a checked (as opposed to unchecked) tone than a syllable with a low register tone T3 or T5 together with the voiced onset /z/ (point estimate = 1.47, z = −4.423, p = 9.7 × 10^{-6}; note that negative z- or t-values are based on the given contrast coding scheme, our report being adjusted for the sake of interpretation).

The three factors of main concern all turned out to have significant effects. Phonation type is crucial to perception: stimuli with tense phonation have a 23.42 to 35.79 times greater odds of being perceived as checked (rather than unchecked) tones than those with modal phonation (point estimate = 28.95, z = +31.098, p = 2.5 × 10^{-212}). Vowel quality also biases perception: stimuli with more centralized /o/ and /ɐ/ have a 7.59 to 11.51 times greater odds of being perceived as checked tones than stimuli with more peripheral /ɔ/ and /ɑ/ (point estimate = 9.35, z = +21.014, p = 4.9 × 10^{-98}). A main effect of the linear trend of duration was revealed: stimuli with the shortest duration have an odds between 2.78 to 4.89 times greater of being identified as checked tones than those with the longest duration (point estimate = 3.68, z = −9.032, p = 1.7 × 10^{-19}).

A negative interaction between phonation type and vowel quality was detected (see Figure 3): the odds of a stimulus with inconsistent phonation and vowel quality cues (i.e., the average of tense phonation + unchecked vowels and modal phonation + checked vowels) identified as a checked tone is between 8.03 and 18.39 times greater than the odds of one with consistent phonation and vowel quality (the average of tense phonation + checked vowels and modal phonation + unchecked vowels) (point estimate = 12.15, z = −11.816, p = 3.2 × 10^{-32}).

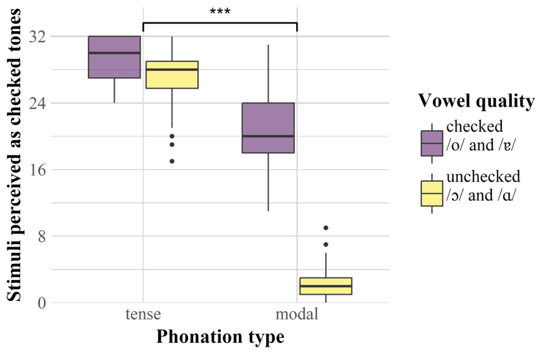

Figure 3: *Distribution of responses across vowel quality and phonation type (32 is the maximum).*

3.2. Reaction time (RT)

A linear mixed-effects regression (*lmer*) model was constructed to analyze the 4630 values of RT under 5125 ms. The fixed and random effects and their contrast coding were the same as in the model reported in Section 3.1 (no random slopes either).

All five main effects and the two-way interaction found in the model on decisions were also significant with respect to RT. Decisions are made between 277 and 375 ms faster on stimuli with vowel /o/ or /ɔ/ than with vowel /ɐ/ or /ɑ/ (point estimate = 326 ms, t[4571.02] = +12.997, p = 5.9 × 10^{-38}). Tone register also makes a difference in the sense that stimuli with T3 or T5 are estimated to take 49 to 148 ms longer to respond to than stimuli with T2 or T4 (point estimate = 99 ms, t[4571.00] = +3.933, p = 8.5 × 10^{-5}).

Tenser phonation takes native listeners between 27 and 125 ms longer to make their decision than modal phonation (point estimate = 76 ms, t[4570.95] = +3.031, p = .0024). Similarly, the more centralized checked vowels /o/ and /ɐ/ turn out to take 24 to 122 ms longer to respond to than the unchecked vowels

/ɔ/ and /ɑ/ (point estimate = 73 ms, t[4571.08] = +2.922, p = .0035). And the longer the duration of a stimulus, the longer the RT turns out to be: from the shortest to the longest stimulus duration, the RT is estimated to rise by 112 to 244 ms (point estimate = 178 ms, t[4571.19] = +5.282, $p = 1.3 \times 10^{-7}$).

The RT difference between checked and unchecked vowels is 328 to 525 ms larger in stimuli with modal than tense phonation (point estimate = 426 ms, t[4571.83] = −8.491, $p = 2.7 \times 10^{-17}$). The shorter-RT advantage of unchecked vowels gets weakened when phonation is tense. In other words, listeners hesitate when auditory information on vowel quality and phonation type conflicts.

For RT we also found two interactions that were not shown in the analysis of decisions, namely duration × phonation type and duration × vowel quality. The RT difference between the two phonation types enlarges by 85 to 350 ms as duration goes from the shortest to the longest (point estimate = 217 ms, t[4570.96] = +3.228, p = .0013). The shorter-RT advantage of unchecked vowels over checked vowels grows by 60 to 324 ms as duration escalates from the shortest to the longest (point estimate = 192 ms, t[4571.06] = +2.843, p = .0045). We found no three-way interaction.

4. Discussion

The current study investigated the perceptual cues to checked tones in Shanghai Chinese and found that all three cues in production (i.e., voice quality, vowel quality and duration) also have an effect on perception. Apart from that, the two conditional factors in our experiment, vowel pairs (/ɐ/–/ɑ/ vs. /o/–/ɔ/) and tone registers (T2 and T4 vs. T3 and T5) also make significant differences with respect to both decisions and RTs.

Tenser phonation, a more central vowel space and shorter duration are signals of checked tones. Not separating glottal constriction and $f0$ in our design, we could only argue that voice quality as a whole plays a crucial role in perception. Tense phonation biases syllables towards T4 and T5 with a longer decision time. In a similar vein, vowel quality has a main effect in checked tone perception: more centrally distributed checked vowels bias listeners towards perceiving a checked tone, but also take them longer to make a decision compared to more peripheral unchecked vowels. Both phonation type and vowel quality cues to checked tones seem to require longer processing time. Nonetheless, this difference in RT is difficult to interpret in the current study. As expected, duration indeed biases the perception of checkedness with a linear trend: the shorter the rhyme, the more likely the syllable is perceived to have a checked tone. Our results also demonstrate that the shorter duration of rhymes gives rise to shorter RTs.

With respect to interactions, the one between phonation type and vowel quality is seen in both decision and RT. When vowel quality and phonation type provide contradictory information on whether a syllable has a checked or unchecked tone, it takes native listeners much longer to make the decision. Yet an average syllable with two conflicting auditory cues (i.e., checked vowel + modal phonation and unchecked vowel + tense phonation lumped together) has a higher chance to be perceived as a checked tone than an average syllable with two consistent cues (i.e., checked vowel + tense phonation and unchecked vowel + modal phonation lumped together).

Two more two-way interactions were observed in the model on RT, one between duration and vowel quality and the other between duration and phonation type, with a common pattern: the shorter-RT advantages of unchecked over checked vowels and of modal over tense phonation enlarge with the increase of duration. Note that the current study did not find evidence that these two interactions involving the effect of duration on RT would influence decision making, and that these findings are based on a monosyllabic word identification task. Duration may function differently on the sentence level.

Which of the three auditory cues to checked tone (tenser phonation, centralized vowel quality, and shorter duration) could be the strongest? When looking at the 95% confidence intervals of the odds ratios of these three main effects, we see that the interval for phonation type runs from 23.42 to 35.79, whereas the interval for vowel quality runs from 7.59 to 11.51. These intervals don't overlap as the upper boundary for vowel quality does not reach the lower boundary for phonation type, which means that for the current stimuli the phonation type is a stronger cue than vowel quality (note that the vowel quality differences, namely in /o/ versus /ɔ/, as well as in /ɐ/ versus /ɑ/, were rather small, so the result might be different for other vowel pairs). The interval for duration runs from 2.78 to 4.89 and doesn't overlap with either of the other two, indicating that duration is an even smaller cue than vowel quality. We can therefore tentatively conclude that tenser phonation seems to be the primary cue in the perception of checked tones in Shanghai Chinese monosyllabic words and that vowel quality and duration are secondary cues. This is generally in line with findings that phonation is primary while duration is secondary to tone perception in other tonal languages with complex tonal systems as Vietnamese [3], [6] and Cantonese [4], and also in line with a recent production study which showed that checked tones in Shanghai Chinese are acoustically realized with tenser phonation and shorter duration of the rhyme [17].

Interestingly, we also found significant effects of the two factors that weren't a priori regarded as cues. Syllables with a low vowel /ɐ/ or /ɑ/ also bring about more checked-tone responses and longer RTs than those with a middle vowel /o/ or /ɔ/, like the effect exhibited by checked vowels that are more centralized. The main effect of this factor on RT implies that the low vowel pair seems more confusing than the middle vowel pair for native listeners to distinguish. A plausible explanation might be that the vowel quality difference between /ɐ/ and /ɑ/ is smaller than that between /o/ and /ɔ/ in terms of F1 as shown in [18], hence longer processing time is required. However, this needs to be further confirmed with more possible checked–unchecked vowel pairs.

Finally, high tones are more likely and more quickly than low tones to be perceived as checked tones. This could be due to any of the various phonetic dimensions that distinguish the two registers. First, high tones have higher $f0$ onsets and longer consonant onsets than low tones (also in this study). Second, there is slight breathiness at the beginning of vowels with low tones T3 and T5 [13], [14], albeit fading away gradually among the young generation [14], [15]. Unfortunately, since this study did not include T1, which also belongs to the high register, it remains unclear whether it is the $f0$ contour of T2 or other features shared across the high register that cause the bias towards perceiving this register as checked tones. It is therefore difficult to draw a conclusion on which factors result in the decision and RT differences between the two registers.

5. Acknowledgements

We would like to express our gratitude to Dirk Jan Vet for setting up the online experiment, and to our participants.

6. References

[1] M. K. Gordon and P. Ladefoged, "Phonation types: A cross-linguistic overview," *Journal of Phonetics*, vol. 29, no. 4, pp. 383–406, Oct. 2001, doi: 10.1006/jpho.2001.0147.

[2] J. E. Andruski, "Tone clarity in mixed pitch/phonation-type tones," *Journal of Phonetics*, vol. 34, no. 3, pp. 388–404, Jul. 2006, doi: 10.1016/j.wocn.2005.07.001.

[3] M. Brunelle, "Tone perception in Northern and Southern Vietnamese," *Journal of Phonetics*, vol. 37, no. 1, pp. 79–96, Jan. 2009, doi: 10.1016/j.wocn.2008.09.003.

[4] K. M. Yu and H. W. Lam, "The role of creaky voice in Cantonese tonal perception," *Journal of the Acoustical Society of America*, vol. 136, no. 3, pp. 1320–1323, 2014, doi: 10.1121/1.4887462.

[5] N. V. Loi and J. A. Edmondson, "Tones and voice quality in modern northern Vietnamese: Instrumental case studies," *Mon-Khmer Studies: A Journal of Southeast Languages*, vol. 28, no. 35, pp. 1–18, 1997.

[6] T. P. Vũ, "The acoustic and perceptual nature of tone in Vietnamese," Ph.D. dissertation, Australian National University, Canberra, Australia, 1981. doi: 10.25911/5d738ecde9342.

[7] R. S. Bauer and P. Benedict, *Modern Cantonese Phonology*. Mouton de Gruyter, 1997.

[8] B. Xu and Z. Tang, *上海市区方言志 [A Description of Shanghainese Spoken in the Urban Districts of the Shanghai City]*. Shanghai: Shanghai Education Press, 1988.

[9] Y. Chen and C. Gussenhoven, "Shanghai Chinese," *Journal of the International Phonetic Association*, vol. 45, no. 3, pp. 321–337, Dec. 2015, doi: 10.1017/S0025100315000043.

[10] Y. Chen, "How does phonology guide phonetics in segment–f0 interaction?," *Journal of Phonetics*, vol. 39, no. 4, pp. 612–625, Oct. 2011, doi: 10.1016/j.wocn.2011.04.001.

[11] N. Ren, "Phonation types and consonant distinctions: Shanghai Chinese," Ph.D. dissertation, The University of Connecticut, Connecticut, US, 1992.

[12] Z. Shen, C. Wooters, and W. S.-Y. Wang, "Closure duration in the classification of stops: A statistical analysis," in *A Festschrift for Ilse Lehiste. Ohio State University Working Papers in Linguistics*, vol. 35, B. D. Joseph and A. M. Zwicky, Eds., 1987, pp. 197–209.

[13] J. Cao and I. Maddieson, "An exploration of phonation types in Wu dialects of Chinese," *Journal of Phonetics*, vol. 20, no. 1, pp. 77–92, Jan. 1992, doi: 10.1016/S0095-4470(19)30255-4.

[14] J. Gao and P. Hallé, "Phonetic and phonological properties of tones in Shanghai Chinese," *Cah Linguist Asie Orient*, vol. 46, no. 1, pp. 1–31, 2017, doi: 10.1163/19606028-04601001.

[15] J. Gao, "Sociolinguistic motivations in sound change: On-going loss of low tone breathy voice in Shanghai Chinese," *Papers in Historical Phonology*, vol. 1, pp. 166–186, Nov. 2016, doi: 10.2218/pihph.1.2016.1698.

[16] J. Edkins, *A Grammar of Colloquial Chinese as Exhibited in the Shanghai Dialect*. Shanghai: Presbyterian Mission Press, 1853.

[17] X. Gao and J. Kuang, "Phonation variation as a function of checked syllables and prosodic boundaries," *Languages*, vol. 7, no. 3, p. 171, Jul. 2022, doi: 10.3390/languages7030171.

[18] Y. Chen, "The acoustic realization of vowels of Shanghai Chinese," *Journal of Phonetics*, vol. 36, no. 4, pp. 629–648, Oct. 2008, doi: 10.1016/j.wocn.2008.03.001.

[19] Y. Chen, "Revisiting the phonetics and phonology of Shanghai tone sandhi," in *Speech Prosody 2008*, Campinas, Brazil: ISCA, May 2008, pp. 253–256. doi: 10.21437/SpeechProsody.2008-55.

[20] M. J. W. Yip, "The tonal phonology of Chinese," Ph.D. dissertation, Massachusetts Institute of Technology, Massachusetts, US, 1980.

[21] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer." 2023. [Online]. Available: https://www.fon.hum.uva.nl/praat/

[22] R Core Team, "R: A language and environment for statistical computing." 2023. [Online]. Available: http://www.r-project.org/

[23] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "lmerTest package: Tests in linear mixed effects models," *Journal of Statistical Software*, vol. 82, no. 13, pp. 1–26, 2017, doi: 10.18637/jss.v082.i13.

[24] A. Zeileis and T. Hothorn, "Diagnostic checking in regression relationships," *R News*, vol. 2, no. 3, pp. 7–10, 2002.