Open Library of Humanities

The production and perception of low tone alternations in Huaiyuan Chinese

Jingfu Zhao, Department of Chinese Language and Literature, Fudan University, Shanghai, China, 21210110027@m.fudan.edu.cn

Hanbo Yan, School of Chinese Studies and Exchange, Shanghai International Studies University, Shanghai, China

Yu-Fu Chien*, Department of Chinese Language and Literature, Fudan University, Shanghai, China, chien_yc@fudan.edu.cn

*Corresponding author.

Huaiyuan Mandarin has three low tone sandhi rules. T1 (low-falling) sandhi and T3 (low-dipping) sandhi involve changing the first low tone to a mid-rising tone when two low tones occur consecutively, potentially yielding neutralization between the sandhi tones and T2 (mid-rising). Half-third sandhi involves abridging the rising portion of T3 before non-low tones, potentially producing neutralization between half-T3 and T1. This study investigated the tonal neutralization between Huaiyuan sandhi tones and their corresponding non-sandhi tones in disyllabic words. For each tone sandhi, 10 sandhi words and 10 non-sandhi words differing only in the first underlying tones were compared. Acoustic results showed nonneutralization for all tone pairs (sandhi-T1/T2, sandhi-T3/T2, half-T3/T1). Identification results revealed complete neutralization between sandhi-T3 and T2, but nonneutralization for the other pairs. Discrimination results demonstrated that native listeners outperformed naïve Huaiyuan listeners in differentiating between half-T3 and T1, suggesting native phonological knowledge helped the perception of sandhi words. These results indicated Huaiyuan low-tone sandhis generally produce noncategorical tone changes. T1 sandhi and T3 sandhi likely stem from phonological dissimilation, while it is hard to justify that half-third sandhi is phonological. These findings contribute to understanding tone sandhi mechanisms across Mandarin dialects.

1. Introduction

Huaiyuan Mandarin is a Chinese dialect spoken in the Anhui Province of China. It belongs to the Jianghuai Mandarin family, so it overlaps greatly in segmental information with Standard Mandarin, or Putonghua (i.e., the official language of China largely based on Beijing Mandarin; Norman, 1988; Chen, 1996; Luo, 2016; Zhang, 2014), but differs strongly from Standard Mandarin in tonal information. Just as Standard Mandarin, Huaiyuan Mandarin is a tone language which uses pitch variation at the syllable level to distinguish word meaning. Huaiyuan has four smooth tones, which are carried by open syllables or syllables closed with a nasal coda, and one checked tone, which is carried by syllables closed with a glottal stop (Gong, 2004; Geng, 2007; Shi, 2007; Zhang & Gong, 2010). Since previous studies investigating the tonal inventory of Huaiyuan mainly employed the impressionistic approach (Gong, 2004; Geng, 2007; Shi, 2007; Zhang & Gong, 2010), there were discrepancies in the exact tone values of Huaiyuan tones (see **Table 1**). Unlike these studies, Zhao (2024) conducted acoustic analysis on the four smooth tones in Huaiyuan. Using a five-point scale notation system in which 5 and 1 represent the highest and lowest pitch within a speaker's pitch range (Chao, 1930, 1968), Zhao (2024) denoted Huaiyuan Tone 1 (e.g., 披 [pʰi] "to drape over"), Tone 2 (e.g., 皮 [pʰi] "skin"), Tone 3 (e.g., 比 [pi] "to compare"), and Tone 4 (e.g., 闭 [pi] "to close") with tonal values of 311, 34, 112, and 51, respectively (see **Figure 1**). Despite the differences in the exact tone values in the literature, these studies generally agreed on the overall height and contour of the four smooth tones. The current study uses mid-low (ML), mid-high (MH), mid-low-mid (MLM), and high-low (HL) to describe the tone height and direction of Huaiyuan T1, T2, T3, and T4.

Interestingly, the two low tones, T1 and T3, in Huaiyuan can undergo their respective tonal alternations, or tone sandhi rules. T1 sandhi refers to the tonal alternation phenomenon in which the first T1 (ML) changes to T2, a mid-rising tone (MH), when followed by another T1, while T3 (MLM) sandhi depicts the tonal alternation phenomenon in which the first T3 (MLM) is

Study	T1	T2	T3	T4
Gong (2003)	21	24	212	51
Geng (2007)	424	45	213	52
Shi (2007)	31	24	224	53
Zhang & Gong (2010)	21	35	212	53
Zhao (2024)	311	34	112	51

Table 1: The smooth tone values of Huaiyuan Chinese, described using the five-point scale notation system developed by Chao (1930, 1968).
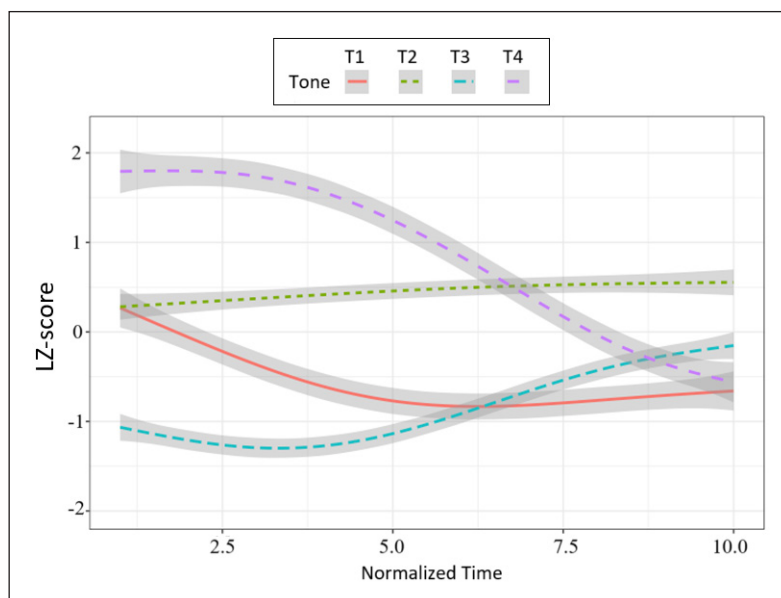
Figure 1: Pitch tracks of the four tones in monosyllables in Huaiyuan Mandarin. The pitch tracks are smoothed. The gray bands represent the confidence interval for a level of .95.

converted to a mid-rising tone (MH) when followed by another T3, as given in (1) below (Gong, 2004). These two sandhi rules can be classified into one type, which involves changing a low tone to a rising tone. They bear resemblance to the well-known T3 sandhi in Standard Mandarin, where the first low-dipping T3 changes to a high-rising tone when followed by another T3 (Chao, 1968; Shih, 1997; Chen, 2000; Lin, 2007; Zhang & Lai, 2010). These sandhi rules in Huaiyuan can potentially be neutralizing, which is defined as the process where phonologically different sounds are treated as phonetically identical (Warner et al., 2004; Dmitrieva et al., 2010; Herd et al., 2010), such that sandhi T1 (ST1) and T2 can be completely neutralized, and so can sandhi T3 (ST3) and T2. The potential complete neutralization suggests that Huaiyuan T1 sandhi and T3 sandhi may yield categorical tone changes since the base T1 and base T3 may be categorically changed to an existing T2 in the tonal inventory of Huaiyuan.

Besides T1 sandhi and T3 sandhi, Huaiyuan T3 also undergoes half-third (HT3) sandhi when it is followed by a non-low tone, where the rising portion of the T3 (MLM) is truncated, resulting in a mid-falling tone (ML), as shown in (1) below (Gong, 2004). This sandhi can be classified into another type, which involves abridging a longer tone to a shorter tone, and is similar to the half-third sandhi in Standard Mandarin in which the low-dipping T3 becomes a low-falling tone (Zhang & Lai, 2010). Unlike Standard Mandarin, the half-third sandhi in Huaiyuan can potentially give rise to tonal neutralization between HT3 (ML) and T1 (ML), therefore being deemed a categorical tone changing rule, while the half-third sandhi in Standard Mandarin cannot, since Standard Mandarin does not have a low-falling tone in its tonal inventory. Given that very little

empirical research has been conducted on the tone sandhis of Mandarin dialects other than Standard Mandarin, and that both of its low tones can undergo tone sandhi, the current study investigated the production and perception of Huaiyuan low tone sandhis by testing the potential acoustic and perceptual neutralization between ST1 and T2, between ST3 and T2, and between HT3 and T1. Results would allow us to understand the acoustic characteristics and perceptual mechanisms of the low tone sandhis in Huaiyuan Mandarin.

(1)   Traditional descriptions of tone sandhi in Huaiyuan[1]
         a.   T1 + T1 → T2 + T1:   ML + ML → MH + ML
         b.   T3 + T3 → T2 + T3:   MLM + MLM → MH + MLM
         c.   T3 + T2/T4 → HT3 + T2/T4:   MLM + MH/HL → ML + MH/HL

1.1. Tone sandhi in Mandarin

Previous tone sandhi research has largely focused on the T3 sandhi and half-third sandhi in Standard Mandarin, and different mechanisms have been proposed for them. Some researchers considered that both sandhi rules are motivated by ease of articulation. For example, Li (2004) proposed that contour tones tend to undergo tonal alternations when concatenated together, so that the articulatory efforts of producing the tone sequence can be reduced. The alternation can be achieved either through contour flattening, such as T3 sandhi, where the dip of T3 is reduced, or through contour removal, such as half-third sandhi, where a U-shaped T3 is converted to a linear-shaped half T3 (HT3). Liu (2004) had a similar viewpoint, suggesting that converting a low-dipping T3 to a rising tone (T3 sandhi) or to a low-falling tone (half-third sandhi) decreases the number of pitch curves in a disyllabic sequence, reducing articulatory efforts. Both T3 and half-third sandhi in Standard Mandarin can be explained by a similar phonetic account.

By contrast, other researchers considered the mechanisms of the two sandhi rules in Standard Mandarin to be different. For example, Liu (1994) proposed that T3 sandhi is a categorical changing rule; namely, after the application of this rule, ST3 loses its contrast with T2 (i.e., both are high-rising). Half-third sandhi is a noncategorical changing rule; therefore, after the application of this rule, HT3 and all the other tones are still contrastive. Shen (1992) as well as Wu (2002) suggested that T3 sandhi is a phonological rule, while half-third sandhi is due to coarticulation occurring at the phonetic level. Wu (2002) further indicated that T3 sandhi is achieved through regressive dissimilation at the phonological level, while half-third sandhi is realized through assimilation at the phonetic level, flattening the tonal contour and reducing the pitch curvature. Zhang and Lai (2010) claimed that half-third sandhi has stronger phonetic motivations, while T3 sandhi has stronger phonological motivations. Since the non-final position

---

[1] In traditional description, the sandhi tone of T1 sandhi is transcribed as 24, and that of T3 sandhi is 35. Both ST1 and ST3 are mid-rising. Meanwhile, although previous studies reported varying pitch values for T2, they consistently characterized it as a mid-rising tone. Hence, the question arises as to whether ST1 and T2, ST3 and T2 are acoustically and perceptually neutralized. We refer to ST1 and ST3 as T2 here for convenience.

prefers a shorter tone, changing a low-dipping T3 to a rising ST3 does not reduce pitch contours on syllables with insufficient duration, whereas, abridging the rising portion of T3 makes the shorter HT3 fit well in this position. Moreover, since plenty of studies have shown that T3 sandhi leads to perceptual neutralization between ST3 and T2, the perceptual distance between the base T3 and the ST3 (T2) is larger than that between the base T3 and the HT3, suggesting stronger phonetic motivations for half-third sandhi than for T3 sandhi (Zhang & Lai, 2010).

From a diachronic point of view, Wang's Qülü, a rhyming dictionary for classical poem composition written between 1573 and 1620, already documented T3 sandhi in Mandarin. Mei (1977), conducting historical linguistic analysis on the tone values of Mandarin in the 16th century, proposed that T3 was low-level, instead of low-dipping, and T2 was low-rising. Given that Mandarin T3 was a low-level tone in the 16th century, there was no motivation for half-third sandhi in that period of time. These pieces of evidence suggest that the two sandhis in Modern Standard Mandarin originated from different sources (Mei, 1977; Zhang & Lai, 2010; Wu, 2020).

In addition to the research about the mechanisms of the two sandhis in Standard Mandarin, another line of research concerns the tonal neutralization between ST3 and T2. This line of studies would help understand whether T3 sandhi is a categorical changing rule. In general, previous acoustic studies did not observe complete neutralization between ST3 and T2 (Zee, 1980; Xu, 1997; Peng, 2000; Yuan & Chen, 2014; Wang et al., 2018; Tu & Chien, 2022). Zee (1980) found that ST3 has a lower turning point and lower offset pitch relative to T2; Peng (2000) showed that ST3 is lower in average pitch height than T2. Yuan and Chen (2014) performed a corpus analysis on telephone speech, finding a later turning point and a smaller pitch rise for ST3 than for T2. They further indicated that the acoustic differences are greater for high frequency words than for low frequency words. Based on the data, Yuan and Chen (2014) proposed that T3 sandhi does not change T3 into T2, dubbing the sandhi tone T3V (i.e., T3 variant). Wang et al. (2018) observed larger acoustic differences between ST3 and T2 when the bearing syllable is stressed, or when the sandhi syllable and its following syllable have a looser rhythmic relationship. These results, together, suggest that although T3 sandhi converts low-dipping T3 to high-rising ST3, ST3 is not acoustically neutralized with T2, showing traces of acoustic characteristics of canonical T3. Therefore, T3 sandhi does not lead to categorical tone change under the scrutiny of acoustic analysis.

Unlike production studies consistently showing nonneutralization between ST3 and T2, results of previous perception studies are inconclusive. Wang and Li (1967) and Peng (2000) conducted identification experiments, finding that native listeners of Standard Mandarin could not correctly identify the target words of ST3 + T3 and T2 + T3 minimal pairs. Identification performance was at the chance level. Tu and Chien (2022) conducted an eye-tracking experiment using the visual world paradigm, revealing that native listeners of Standard Mandarin were somewhat sensitive to the acoustic details between ST3 and T2, and could use the information for lexical access. Using a more sensitive method, Tu and Chien (2022) observed results inconsistent with most of

the previous studies collecting overt behavioral responses. These perception results showed that despite the acoustic differences between ST3 and T2, native listeners may not always be able to capture the differences unless a more delicate method is employed. Thus, whether T3 sandhi in Standard Mandarin is a categorical tone changing rule in perception is still unclear.

Unlike the above-mentioned studies focusing on Standard Mandarin, Zhang and Liu (2011) as well as Li and Chen (2016) acoustically analyzed the tone sandhi and tonal coarticulation in Tianjin Mandarin, a Mandarin dialect classified under the Jilu Mandarin family. They found that both T1 sandhi and T3 sandhi were nonneutralizing, leading to noncategorical tone changes (i.e., T1 sandhi: T1 (L) + T1 (L) → T3 (LH) + T1 (L); T3 sandhi: T3 (LH) + T3 (LH) → T2 (H) + T3 (LH)). They also confirmed that both rules are sandhis since the F0 realizations of the sandhi tones were unpredictable from their respective canonical tones; thus, they should be phonological. In contrast with Li and Chen (2016), Zhang and Liu (2011) also observed the nonneutralizing half-third sandhi (i.e., T3 (LH) + T4 (HL) → T1 (L) + T4 (HL)). Specifically, they found small pitch rise at the end of HT3, suggesting gradient reactions to insufficient duration for the non-final syllable to manifest a full-rising T3. Hence, they proposed that the half-third sandhi in Tianjin may be rooted in coarticulation. Despite thorough acoustic investigations, no perceptual experiment was conducted, leaving the question of perceptual neutralization in Tianjin tone sandhi unanswered.

In summary, previous studies on tone sandhi have primarily focused on three aspects: whether a sandhi rule results in neutralization, whether it constitutes a categorical change, and whether it operates as a phonological rule. Based on Peng (2000), Tu and Chien (2022), and Yuan and Chen (2014), we define a categorical sandhi rule as one that results in complete neutralization between the sandhi tone and an existing canonical tone within the tonal inventory. Further, following Li and Chen (2016) and Zhang and Liu (2011), we regard a sandhi rule as phonological if the realization of the sandhi tone cannot be phonologically predicted from its base tone. It is critical to note that while complete neutralization is a necessary condition for categoricity, phonological status does not inherently entail categorical tone change, as exemplified by the nonneutralizing yet phonologically motivated T3 sandhi in Standard Mandarin. These criteria address distinct dimensions of sandhi rules. Building on these foundations, the current study investigated the potential acoustic and perceptual neutralization of Huaiyuan's low tone sandhis (T1 sandhi, T3 sandhi, and half-third sandhi) and discussed the phonological predictability of their realizations. By doing so, we aimed to determine whether these sandhi processes constitute categorical changes, phonological rules, or both, thereby contributing to a clearer classification of tone sandhi phenomena in understudied Mandarin dialects.

1.2. The current study

The vast majority of studies have focused on the tone sandhis in Standard Mandarin. Very few studies have examined both the acoustic and perceptual aspects of the neutralization of tone

sandhi in the other Mandarin dialects. Therefore, whether tone sandhis in these Mandarin dialects are neutralizing in production and perception is still unclear. Given that Huaiyuan T1 and T3 are both low tones, and their respective sandhi rules may not result from the same mechanisms as the sandhi rules in the other Mandarin dialects, it is warranted to investigate the acoustic and perceptual neutralization between the sandhi tones and their corresponding canonical tones in Huaiyuan Mandarin, so that the categoricity of these low tone sandhi rules can be further understood and compared cross-linguistically. Unlike Standard Mandarin, Huaiyuan Mandarin offers us an opportunity to compare its T1 sandhi and T3 sandhi, both of which change a low tone to a rising tone. Through this comparison, we would be able to see whether this type of sandhi results from similar mechanisms. Regarding Huaiyuan half-third sandhi, the existence of the low-falling T1 allows us to directly compare the acoustic and perceptual neutralization between Huaiyuan T1 and HT3, shedding light on whether Huaiyuan half-third sandhi can be considered a categorical tone changing rule. Such comparisons are absent in Standard Mandarin due to the lack of a second low-falling tone. Through investigating the acoustic realizations of these sandhi tones, we would also be able to infer whether these low tone sandhis in Huaiyuan are phonological or not.

In order to address these issues, a production experiment, an identification experiment, and a discrimination experiment were conducted. Specific research questions include: 1) whether Huaiyuan ST1 and T2, ST3 and T2, as well as HT3 and T1 are neutralized in production, 2) whether Huaiyuan ST1 and T2, ST3 and T2, as well as HT3 and T1 are neutralized in perception, and 3) whether native Huaiyuan listeners would be able to use the acoustic differences, and their native phonological knowledge, in perceiving the two minimal tone pairs in Huaiyuian. These experiments were reviewed and approved by the Human Subjects Committee of the Department of Chinese Language and Literature at Fudan University.

2. Production experiment

The production experiment aimed to empirically investigate whether Huaiyuan ST1 and T2, ST3 and T2, and HT3 and T1 are acoustically completely neutralized, in order to understand the production characteristics and the categoricity of Huaiyuan low tone sandhis. Pitch height and pitch contour of the two tones within a tone pair were compared. If complete neutralization is observed, it would indicate that the sandhi rule leads to categorical tone change in production. If nonneutralization is observed, it would suggest that the sandhi rule does not result in categorical tone change in production. In addition, whether the F0 realization of the sandhi tone is unpredictable from its canonical tone would be the determinant of whether or not the sandhi rule is phonological. An unpredictable sandhi tone would suggest that the sandhi rule is phonological, while a predictable sandhi tone would suggest that the sandhi rule can be explained by phonetic processes alone (Li & Chen, 2016).

Based on previous production studies in Standard Mandarina and Tianjin Mandarin (Peng, 2000; Yuan & Chen, 2014; Zhang & Liu, 2011; Li & Chen, 2016), we predicted that all the sandhis would lead to nonneutralization in production, revealing traces of the acoustic characteristics of the canonical tones from which the sandhi tones are derived. We also predicted that the tonal contours of ST1 and ST3 would not be predictable from that of T1 and T3, but the tonal contour of HT3 would be similar to the first half of T3. Therefore, although all the three sandhi rules would result in noncategorical tone changes in production, Huaiyuan T1 sandhi and T3 sandhi should be phonological, but Huaiyuan half-third sandhi would not be justified as a phonological rule. The production data collected in this experiment also served as the ground for the two perception experiments.

2.1. Participants

Twelve native speakers of Huaiyuan Chinese (7 males, 5 females), aged between 40 and 60 (M = 44.75), were recruited as participants. In order to minimize the influence of Standard Mandarin, all the participants were born and raised in the city of Huaiyuan, and their parents and spouses were all Huaiyuan locals. They did not have any experience of living in another city besides Huaiyuan before joining this experiment. They all received elementary school education, and some also received middle school education. They had no difficulty reading the Chinese characters on the stimulus list. They did not have any reported language production or comprehension impairment at the time of testing. All the participants were asked to provide informed consent before the production experiment and were paid for their participation.

2.2. Stimuli

Ten minimal pairs of ST1 + T1 and T2 + T1 words (e.g., 巫山 "Mountain Wu" vs. 吴山 "Mountain Wu", [u san]), ten minimal pairs of ST3 + T3 and T2 + T3 words (e.g., 彩礼 "betrothal gifts" vs. 财礼 "monetary price", [tsʰɛ li]), as well as ten minimal pairs of HT3 + T4 and T1 + T4 words (e.g., 企盼 "hope for" vs. 期盼 "expect", [tɕʰi pʰan]) were selected as critical stimuli (see Appendix A). Confirmed by two native Huaiyuan speakers who did not participate in the experiment, all the critical words were common words in Huaiyuan. Moreover, since there is no corpus for Huaiyuan, a corpus for Standard Mandarin was used to estimate the word frequencies of the critical words. Results of t tests showed that ST1 + T1 and T2 + T1 words did not differ in word frequency (t(18) = 1.101, p = .288). Neither did ST3 + T3 and T2 + T3 words (t(18) = .502, p = .622), or HT3 + T4 and T1 + T4 words (t(18) = −.5, p = .623). Lastly, the sandhi sets and their corresponding non-sandhi sets were also matched in stroke number (ST1 + T1 vs. T2 + T1, t(18) = −.065, p = .949; ST3 + T3 vs. T2 + T3, t(18) = −.5, p = .623; HT3 + T4 vs. T1 + T4, t(18) = −1.021, p = .321).

In addition to the critical words, ten T1 + T2 (ML + MH), ten T3 + T2 (MLM + MH), twenty T4 + T2 (HL + MH), and twenty T4 + T4 (HL + HL) filler words were also included to balance the numbers of each tone type of the entire stimulus set. In total, 40 words began with a rising tone; 40 began with a low tone, and 40 began with a high-falling tone. Forty ended with a low tone; 40 ended with a rising tone, and 40 ended with a high-falling tone.

2.3. Procedure

Participants first completed a consent form. Then, they were seated about 70 cm in front of a laptop screen in a quiet room, and were recorded using the Adobe Audition software, with a cardioid microphone (Shure, model SM57) and a digital solid-state recorder (Zoom H4N) at a sampling rate of 44,100 Hz and the resolution of 16 bits. The stimuli were presented via Paradigm (Tagliaferri, 2019).

In a trial, a fixation cross was first presented in the middle of the screen for 500 ms, then a word for 2000 ms. Participants were instructed to produce the word as naturally as possible. Five practice trials were first provided to the participants to familiarize them with the experimental procedure. After the practice session, the main experiment commenced. The main experiment consisted of 240 trials, with 60 critical words and 60 filler words randomly presented twice. The whole experiment took approximately 20 minutes.

2.4. Pitch track analysis

Pitch tracks of the first syllables of the critical words were analyzed using Praat (Boersma & Weenink, 2023). For rising tones ST1, ST3, and T2, pitch tracks were measured from the onset of the second periodicity in the waveform to the peak in the pitch track analysis in Praat. For falling tones HT3 and T1, pitch tracks were measured from the onset of the second periodicity in the waveform to the lowest point in the pitch track analysis in Praat (Zhu, 2010). Ten equally spaced measurement points were extracted for each pitch track using the ProsodyPro Praat script (Xu, 2005, 2010). The extracted pitch tracks were checked for octave jumps. When there was an octave jump, the values of the measurement points were manually corrected using $F0 = 1/T(s)$ in which T refers to the duration of one period of the waveform. The values of the measurement points were then log-transformed with base 10. After the process, the log-transformed values were z-transformed using the formula in (2) to minimize pitch variation due to gender and speaker identity (Zhu, 2010). The log-z scores (LZ-scores) were subject to subsequent statistical analysis.

(2) $LZ = (x_i - \mu)/\sigma,$

where x_i refers to the log-transformed value of each measurement point, μ refers to the mean of the overall log-transformed values of a given speaker, and σ represents the SD of the overall log-transformed values of a given speaker.

2.5. Production results

Growth curve analysis (Mirman, 2014) was conducted on the LZ-scores to model the pitch tracks of ST1 and T2, the pitch tracks of ST3 and T2, as well as the pitch tracks of HT3 and T1 using the lme4 package in R (Bates et al., 2015), with p-values calculated by the lmerTest package (Kuznetsova et al., 2017). Considering the possible shape of Chinese tones (Tu & Chien, 2022), the linear, quadratic, and cubic time polynomials, which represent linear, U shape, and S shape of the tones, were included one at a time as fixed factors to estimate the overall tonal contour of each pair of tones. Participants and items were included as random intercepts for the random effects structure. Then, the linear, quadratic, and cubic time polynomials were entered one at a time as by-participant and by-item random slopes into the random effects structure. Whenever a model could not converge or was singular, a simpler random effects structure was used (Li, 2023). Through this process, three models for each tone pair were generated, with one, two, or three time polynomials as fixed factors, and then compared using likelihood ratio tests to determine the optimal basic model to which subsequent fixed factors were added. For all three tone pairs, the basic model contained all three time polynomials (see **Table 2**).

ST1/T2 pair	χ^2	df	p
Linear vs. quadratic	31.845	1	<.001
Quadratic vs. cubic	16.505	1	<.001
Best basic model: LZ-scores ~ (ot1 + ot2 + ot3) + (1 + ot1 | Participant) + (1 + ot1 | item)			
ST3/T2 pair	χ^2	df	p
Linear vs. quadratic	47.437	4	<.001
Quadratic vs. cubic	14.274	1	<.001
Best basic model: LZ-scores ~ (ot1 + ot2 + ot3) + (1 + ot1 + ot2 | Participant) + (1 + ot1 | item)			
HT3/T1 pair	χ^2	df	p
Linear vs. quadratic	27.485	4	<.001
Quadratic vs. cubic	7.824	1	.005
Best basic model: LZ-scores ~ (ot1 + ot2 + ot3) + (1 + ot1 + ot2 | Participant) + (1 + ot1 | item)			

Table 2: Likelihood ratio tests on the number of time polynomials as fixed effects for each tone pair.

In order to further evaluate whether the two tones of each pair were completely neutralized, for each tone pair, two additional models were built based on the basic model, with TONEPAIR (sandhi vs. canonical) and interactions between TONEPAIR and time polynomials, added one at a time as fixed factors. Likelihood ratio tests were employed to determine the optimal model.

Table 3 shows that for ST1 and T2, Model2 was significantly better than Model1 and the basic model; therefore, it was chosen as the best model. Within Model2, the linear, quadratic, and cubic time polynomials were all significant, indicating that the overall shape of ST1 and T2 was an incomplete S-shape on an angle as compared to a horizontal line. In addition, TONEPAIR was significant (ST1 baseline). The positive estimate revealed that T2 was significantly higher than ST1 in average pitch height. The interaction between TONEPAIR and the linear time term was also significant. The negative estimate demonstrated that T2 had a more negative slope relative to ST1. Lastly, the interaction between TONEPAIR and the quadratic time term was also significant. The negative estimate indicated that T2 had a less convex shape (U-shape) than ST1. The pitch tracks of ST1 and T2 are shown in **Figure 2**.

ST1/T2 pair	χ^2	df	p
Basic vs. M1	32.780	1	<.001
M1 vs. M2	8.806	3	.032
Basic model: LZ-scores ~ (ot1 + ot2 + ot3) + (1 + ot1 | Participant) + (1 + ot1 | item) Model1: LZ-scores ~ (ot1 + ot2 + ot3) + TONEPAIR + (1 + ot1 | Participant) + (1 + ot1 | item) ✔ Model2: LZ-scores ~ (ot1 + ot2 + ot3) * TONEPAIR + (1 + ot1 | Participant) + (1 + ot1 | item)			
Model2	β	SE	t	p
(Intercept)	−.236	.074	−3.165	.003
Linear	1.075	.105	10.258	<.001
Quadratic	.233	.043	5.393	<.001
Cubic	−.128	.043	−2.967	.003
TONEPAIR	.634	.079	8.065	<.001
Linear:TONEPAIR	−.198	.088	−2.238	.036
Quadratic:TONEPAIR	−.123	.060	−2.050	.040
Cubic:TONEPAIR	.013	.060	.224	.823

Table 3: Likelihood ratio tests on fixed effects for the ST1/T2 pair and the results of fixed effects estimation for the best model.

Table 4 shows that for ST3 and T2, Model4 could not significantly improve the fit over Model3, and Model3 was significantly better than the basic model. Thus, Model3 was selected as the optimal model. Within this model, the linear and cubic time polynomials were significant, suggesting that the overall shape of ST3 and T2 was an S-shape on an angle relative to a horizontal line. TonePair (ST3 baseline) was also significant. The negative estimate revealed that the average pitch height of ST3 was significantly lower than that of T2. The pitch tracks of ST3 and T2 are shown in **Figure 2**.

ST3/T2 pair	χ^2	df	p
Basic vs. M3	4.552	1	.033
M3 vs. M4	1.042	3	.791
Basic model: LZ-scores ~ (ot1 + ot2 + ot3) + (1 + ot1 + ot2 | Participant) + (1 + ot1 | item) ✔ Model3: LZ-scores ~ (ot1 + ot2 + ot3) + TonePair + (1 + ot1 + ot2 | Participant) + (1 + ot1 | item) Model4: LZ-scores ~ (ot1 + ot2 + ot3) * TonePair + (1 + ot1 + ot2 | Participant) + (1 + ot1 | item)			

Model3	β	SE	t	p
(Intercept)	.612	.082	7.471	<.001
Linear	.879	.146	6.024	<.001
Quadratic	.152	.075	2.029	.066
Cubic	−.128	.034	−3.781	<.001
TonePair	−.151	.066	−2.279	.032

Table 4: Likelihood ratio tests on fixed effects for the ST3/T2 pair and the results of fixed effects estimation for the best model.

Table 5 reveals that for HT3 and T1, Model6 was significantly better than Model5 and the basic model. Hence, it was chosen as the best mode. Within Model6, the linear and cubic time polynomials reached significance, and the quadratic time term showed a trend towards significance, meaning that the overall shape of HT3 and T1 was an incomplete S-shape on an angle as compared to a horizontal line. Besides, TonePair (T1 baseline) was significant. The negative estimate revealed that the average pitch height of HT3 was significantly lower than that of T1. The interaction between the linear time term and TonePair was also significant. The positive estimate displayed that HT3 had a more positive slope than T1, meaning that the decline

of HT3 was not as drastic as that of T1. Moreover, the interaction between the quadratic time term and TonePair was significant. The positive estimate means that HT3 was more U-shaped than T1. Finally, the interaction between the cubic time term and TonePair reached significance. The negative estimate indicated that HT3 was less S-shaped than T1. The pitch tracks of HT3 and T1 are shown in **Figure 2**.

HT3/T1 pair	χ^2	*df*	*p*
Basic vs. M5	20.034	1	<.001
M5 vs. M6	24.303	3	<.001
Basic model: LZ-scores ~ (ot1 + ot2 + ot3) + (1 + ot1 + ot2 | Participant) + (1 + ot1 | item) Model5: LZ-scores ~ (ot1 + ot2 + ot3) + TonePair + (1 + ot1 + ot2 | Participant) + (1 + ot1 | item) ✔ Model6: LZ-scores ~ (ot1 + ot2 + ot3) * TonePair + (1 + ot1 + ot2 | Participant) + (1 + ot1 | item)			
Model6	β	*SE*	*t*	*p*
(Intercept)	−.582	.074	−7.851	<.001
Linear	−1.944	.193	−10.065	<.001
Quadratic	.088	.050	1.758	.088
Cubic	.151	.045	3.398	<.001
TonePair	−.581	.080	−7.238	<.001
Linear:TonePair	.790	.163	4.843	<.001
Quadratic:TonePair	.123	.062	1.985	.047
Cubic:TonePair	−.126	.061	−2.050	.040

Table 5: Likelihood ratio tests on fixed effects for the HT3/T1 pair, and the results of fixed effects estimation for the best model.

Taken together, these production results showed that although Huaiyuan participants successfully applied T1 sandhi, T3 sandhi, and half-third sandhi to the critical words, the sandhi tones and their corresponding canonical tones were not neutralized in production. For ST1, it was lower, steeper, and more U-shaped than T2; for ST3, it was lower than T2; for HT3, it was lower and less S-shaped than T1. The initial decline of HT3 was also not as sharp as that of T1. Based on these results, Huaiyuan T1 sandhi, T3 sandhi, and half-third sandhi are nonneutralizing

in production, leading to noncategorical tone changes. It is consistent with the production data obtained in Standard Mandarin and Tianjin Mandarin (Zee, 1980; Xu, 1997; Peng, 2000; Yuan & Chen, 2014; Wang et al., 2018; Tu & Chien, 2022; Zhang & Liu, 2011; Li & Chen, 2016). Furthermore, given that the pitch tracks of both ST1 and ST3 were unpredictable from their corresponding canonical tones, Huaiyuan T1 sandhi and T3 sandhi should be phonological. Conversely, Huaiyuan half-third sandhi can be explained by phonetic processes alone since HT3 appeared to be the first half of its base tone (Li & Chen, 2016).
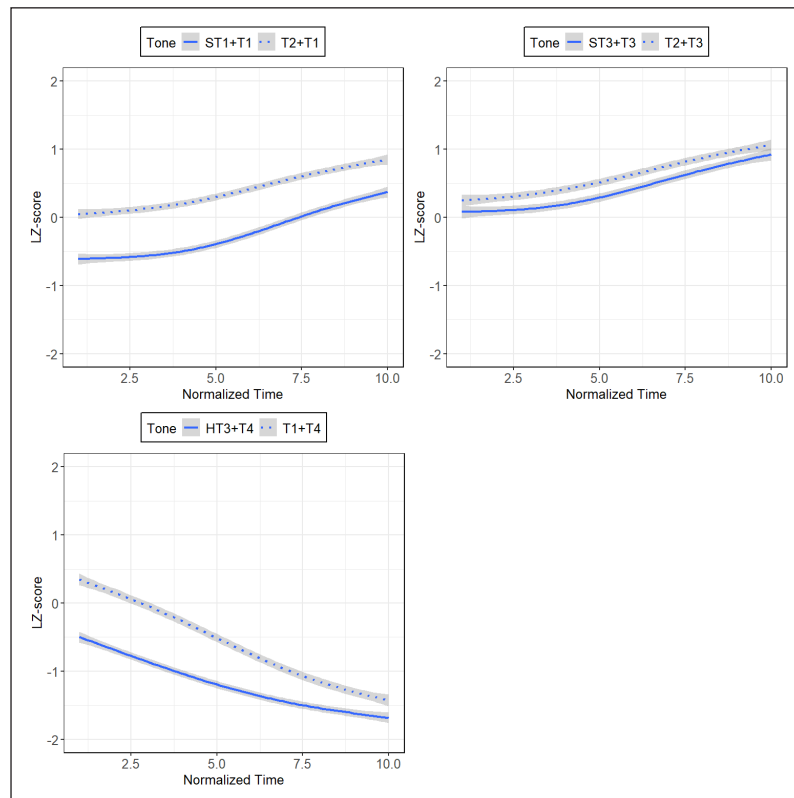

Figure 2: The top-left panel represents the pitch tracks of the first syllables of ST1 + T1 and T2 + T1 words. The top-right panel represents the pitch tracks of the first syllables of ST3 + T3 and T2 + T3 words. The bottom-left panel represents the pitch tracks of the first syllables of HT3 + T4 and T1 + T4 words. The pitch tracks are smoothed. The gray bands represent the confidence interval for a level of .95.

3. Perception experiments

While the production results showed significant acoustic distinctions between ST1 and T2, ST3 and T2, and between HT3 and T1, such acoustic differences between the two tones within a pair do not inherently ensure that they can be used in perception (Peng, 2000). To determine whether these acoustic differences are functionally encoded in listeners' phonological categories,

we first conducted an identification experiment, testing whether Huaiyuan ST1 and T2, ST3 and T2, and HT3 and T1 are perceptually neutralized or remain distinct. Based on the fact that the acoustic differences between ST3 and T2 appeared smaller than those between ST1 and T2, and between HT3 and T1, we predicted that ST3 and T2 may be perceptually neutralized, as observed in the Standard Mandarin study of Peng (2000). In contrast, we predicted that ST1 and T2, as well as HT3 and T1 would maintain perceptual distinctiveness. Such results would indicate that Huaiyuan ST3 may be perceived as T2, which is a categorical tone change, while Huaiyuan T1 sandhi and half-third sandhi are not categorical tone-changing rules. This perception pattern would further demonstrate that although both T1 sandhi and T3 sandhi should be deemed phonological, as evidenced by the fact that the pitch tracks of ST1 and ST3 cannot be inferred from their respective canonical tones, they may exhibit different perceptual characteristics. The integration of the production and perception data would allow us to discern the similarities and differences among Huaiyuan T1 sandhi, T3 sandhi, and half-third sandhi.

In addition to the identification experiment, a discrimination experiment was also conducted to further investigate whether Huaiyuan phonological knowledge was also involved in helping Huaiyuan listeners' sandhi word identification. In order to address this question, native Huaiyuan listeners and naïve Huaiyuan listeners were asked to discriminate between Huaiyuan minimal tone pairs. If Huaiyuan experience improves discrimination performance, it would suggest that native Huaiyuan listeners not only rely on surface acoustic information, but also depend on Huaiyuan phonological knowledge in perception. These two perception experiments were reviewed and approved by the Human Subjects Committee of the Department of Chinese Language and Literature at Fudan University.

3.1. Identification experiment

3.1.1. Participants

Since it was not easy to recruit enough participants for the middle-aged group alone, 20 middle-aged participants (8 males, 12 females), aged between 40 and 60 ($M = 46.55$), and 20 younger participants (11 males, 9 females), aged between 20 and 35 ($M = 26.85$), were recruited. A language background questionnaire was then provided to ensure that they had native Huaiyuan proficiency and had no language production or comprehension impairment at the time of testing. According to their self-reports, they were all native speakers of Huaiyuan. Both of their parents were also native speakers of Huaiyuan. They all grew up in Huaiyuan City and had not lived in a city other than Huaiyuan for more than 5 years. Given that the two groups of participants differed strongly in age, for statistical analysis, AGE was set as an independent variable to evaluate whether it influenced the identification results. All the participants provided informed consent before the main experiment and were paid for their participation after the experiment.

3.1.2. Stimuli

Ten minimal pairs of ST1 + T1 and T2 + T1, ten minimal pairs of ST3 + T3 and T2 + T3, and ten minimal pairs of HT3 + T4 and T1 + T4 used in the production experiment were included in the identification experiment. The critical words produced by Speaker12 and Speaker3 were selected since their pitch tracks were the closest to the average pitch tracks of the critical words produced by all the speakers. Therefore, the productions of these two speakers were the best representatives of the entire set of the critical tone productions.

3.1.3. Procedure

Participants were invited to a quiet room, seated around 70 cm in front of a laptop screen, and wore headphones with the sound volume adjusted to a comfortable level. A two-alternative forced choice identification experiment was run by the PsychoPy software (Peirce et al., 2022). In each trial, a fixation cross appeared in the middle of the screen for 1000 ms. After that, a minimal pair of words (e.g., 巫山 "Mountain Wu", 吴山 "Mountain Wu"; ST1 + T1 vs. T2 + T1; [wu san]) written in Simplified Chinese characters were presented to the right and left sides of the fixation location for 1500 ms, during which participants were asked to look at the words. Then, an auditory stimulus was presented via the headphones, and the participants were instructed to make a behavioral response by pressing the "F" key with the left hand and the "J" key with the right hand on the keyboard. The "F" key referred to the word on the left, and the "J" key referred to the word on the right. Eight practice trials were presented first to ensure that the participants fully understood the task. After that, the main experiment was presented, with the two speakers' voices separated in two blocks. The order of the blocks was counterbalanced across participants, and the stimuli within each block were presented in a randomized order. The main experiment consisted of 120 trials (60 words × 2 speakers). The whole experiment took approximately 20 minutes.

3.1.4. Results

A-prime scores were calculated using the formula in (3) to evaluate participants' identification performance (Grier, 1971; Snodgrass et al., 1985; Peng, 2000; So & Best, 2010). A-prime (A') analysis considers not only correct responses, but also false alarms. Therefore, it is a better measurement to reflect participants' signal detectability. A-prime scores range from 0 to 1, with .5 meaning chance-level performance and 1 representing perfect performance.

(3) $A' = .5 + [(y - x) (1 + y - x) / 4y (1 - x)]$,

where x is the ratio of false alarms (i.e., identification of sandhi words for non-sandhi words and that of non-sandhi words for sandhi words), and y is the ratio of correct responses (i.e., identification of sandhi words for sandhi words and that of non-sandhi words for non-sandhi words).

For the ST1 + T1 and T2 + T1 words, the middle-aged group obtained an average A' score of .563 (*SD* = .125), while the younger group obtained an average A' score of .585 (*SD* = .161). Two one-sample t tests were conducted on the A' scores with a test value of .5 for each age group to evaluate whether participants' responses were significantly better than chance-level responses. Results showed that the average score of the middle-aged group was significantly higher than .5 ($t(19)$ = 2.231, p = .038 < .05), and so was the average score of the younger group ($t(19)$ = 2.302, p = .033 < .05), indicating that ST1 and T2 were not perceptually neutralized (see **Figure 2**).

For the ST3 + T3 and T2 + T3 words, the average A' scores for the middle-aged and younger groups were .512 (*SD* = .1) and .492 (*SD* = .085), respectively. One-sample t tests with A' scores as the dependent variable and a test value of .5 showed that the average score of the middle-age group and that of the younger group were not significantly different from .5 (middle-aged group: $t(19)$ = .194, p = .849; younger group: $t(19)$ = −.335, p = .741), suggesting that ST3 and T2 were perceptually completely neutralized (see **Figure 2**).

For the HT3 + T4 and T1 + T4 words, the average A' scores for the middle-aged and younger groups were .81 (*SD* = .121) and .883 (*SD* = .067), respectively. Two one-sample t tests were performed on participants' A' scores, with a test value of .5. Results showed significant differences between the A' scores and the test value for both age groups (middle aged group: $t(19)$ = 11.534, p < .01; younger group: $t(19)$ = 25.489, p < .01), indicating that HT3 and T1 were not perceptually neutralized (see **Figure 3**).

In order to examine whether the effects of AGE (middle-aged, younger), TONEPAIR (ST1/T2, ST3/T2, HT3/T1) and their interaction significantly influenced participants' A' scores, a linear mixed-effects model was built with AGE, TONEPAIR, and their interaction as fixed factors. Random intercepts as well as by-participant random slopes of AGE were entered into the random effects structure. The analysis was conducted using the lme4 package in R (Bates et al., 2015), with p values calculated by the lmerTest package (Kuznetsova et al., 2017). Results showed that the ST1/T2 and ST3/T2 pairs elicited significantly lower A' scores than the HT3/T1 pair (baseline: HT3/T1; ST1/T2 vs. HT3/T1: β = −.286, *SE* = .033, t = −8.669, p < .001; ST3/T2 vs. HT3/T1: β = −.346, *SE* = .033, t = −10.502, p < .001). The ST3/T2 pair showed a trend towards yielding significantly lower A' scores than the ST1/T2 pair (baseline: ST1/T2; β = −.060, *SE* = .033, t = −1.833, p = .071). AGE and the interaction between AGE and TONEPAIR were not significant (AGE: $F(1)$ = .565, p = .457; AGE × TONEPAIR: $F(2)$ = .493, p = .613), indicating that the two groups of participants produced similar A' scores, and the A' score difference within each tone pair was not modulated by AGE.

These identification results suggested that the participants were sensitive to the acoustic differences between ST1 + T1 and T2 + T1 words, as well as between HT3 + T4 and T1 + T4 words, and were able to use the subtle acoustic differences within minimal pairs for word identification. However, they could not utilize the small, yet statistically significant, acoustic differences

between ST3 + T3 and T2 + T3 words for word identification. The perceptual results of the ST1/T2 and HT3/T1 pairs were consistent with the production results, revealing nonneutralization. However, the perceptual results of the ST3/T2 pair contrasted with the production results, showing complete neutralization. Based on the data, Huaiyuan ST3 may be perceived as T2, which is a categorical tone change, but Huaiyuan T1 sandhi and half-third sandhi are not categorical tone-changing rules. **Table 6** summarizes the results of the production and identification experiments.
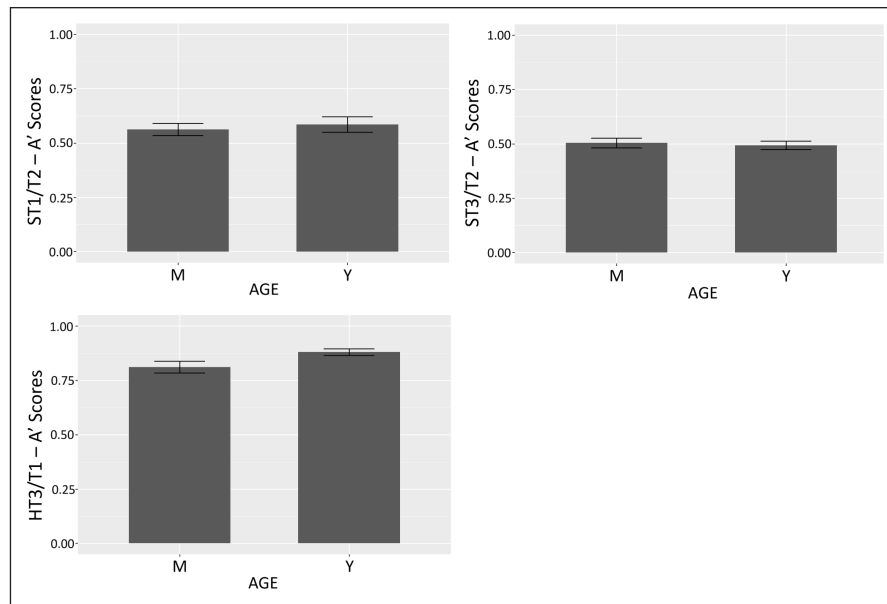

Figure 3: A' scores of the three minimal pairs between the sandhi and non-sandhi words presented as boxplots. M represents the middle-aged group. Y represents the younger group. Error bars are plotted with a confidence level of .95.

Minimal tone pairs	Production results	Identification results
ST1 + T1 vs. T2 + T1	Nonneutralization	Nonneutralization
ST3 + + T3 vs. T2 + T3	Nonneutralization	Complete neutralization
HT3 + T4 vs. T1 + T4	Nonneutralization	Nonneutralization

Table 6: Summary of production and identification results.

3.2. Discrimination experiment

Although the nonneutralizing results of the identification experiment could be attributed to Huaiyuan participants' sensitivity to the surface acoustic differences between the words of a minimal tone pair, they might also be explained by participants' utilization of Huaiyuan phonological knowledge, activating the underlying tones of the sandhi words, which further

facilitated word identification. In order to examine the possibility of using Huaiyuan phonological knowledge in word identification, a discrimination experiment comparing native and naïve Huaiyuan listeners was conducted, thereby evaluating whether native language experience plays a critical role in discriminating Huaiyuan tones. The importance of selecting naïve Huaiyuan listeners who know a different tone language was that they could only rely on their acoustic-phonetic knowledge of non-Huaiyuan tones. If they outperformed or performed comparably to native Huaiyuan listeners, it would reveal that lower-level acoustic-phonetic knowledge or knowledge of non-Huaiyuan tones was sufficient, with language-specific experience being nonessential. Conversely, if their discrimination performance fell short of that demonstrated by native Huaiyuan listeners, it would indicate that native Huaiyuan listeners likely leverage additional phonological knowledge in discriminating these tone pairs. Based on the results of the previous two experiments, we only selected the HT3 + T4 and T1 + T4 words as stimuli since they showed the largest acoustic and perceptual differences among the three groups of words. Using these words could avoid the floor effect and granted us a better chance to observe performance differences between the native and nonnative participants if there were any. We predicted that native Huaiyuan listeners would outperform naïve Huaiyuan listeners in discriminating between Huaiyuan HT3 + T4 and T1 + T4 minimal pairs, suggesting the recruitment of native phonological knowledge in discrimination.

3.2.1. Participants

Twenty native Huaiyuan listeners who had participated in the identification experiment around 5 months ago participated in the discrimination experiment. Twenty naïve Huaiyuan listeners were also recruited (6 males, 14 females; mean age: 25.7). They were all native listeners of a Mandarin dialect other than Jianghuai Mandarin to which Huaiyuan Mandarin belongs. All the participants did not have any language production or comprehension impairment at the time of testing. They all provided informed consent before the experiment, and received monetary compensation after the experiment.

3.2.2. Stimuli

The 10 minimal pairs of the HT3 + T4 and T1 + T4 words used in the production and identification experiments were used in this experiment. The stimuli were all produced by Speaker12 in the production experiment.

3.2.3. Procedure

Participants were invited to a quiet room, seated about 70 cm in front of a laptop screen, and wore headphones with sound volume adjusted to the comfortable level. An experimenter then introduced the procedure to the participants, and started an AX discrimination experiment

implemented in PsychoPy (Peirce et al., 2022). After four practice trials, participants were asked if they had any questions, and then the main experiment was provided. In a trial, a fixation cross first appeared in the middle of the screen for 500 ms. Then, two stimuli were presented one after another with a 500 ms interstimulus interval. Immediately after the disappearance of the second stimulus, the word "相同" (same) and "不同" (different) were presented to the left and right of the previous fixation location, and the participants were instructed to judge whether the two stimuli had the same initial tone as quickly and as accurately as possible, by pressing "F", representing different, and "J", representing the same, on the keyboard. If the participants did not make any response within 5000 ms, this particular trial would be recorded as an incorrect response, and the next trial would begin. The discrimination experiment consisted of 4 practice trials and 80 main trials (10 pairs of words, 4 combinations, AA, BB, AB and BA, for each pair, 2 repetitions). The trial order was fully randomized within one repetition. The whole experiment lasted approximately 20 minutes.

3.2.4. Results

Participants' discrimination performance was evaluated via D-prime (d') analysis (Francis & Ciocca, 2003; Macmillan & Creelman, 1991), which considers both sensitivity and bias. Participant's d' scores were calculated using the formula in (4) below. The logic and pragmatic range of d' is from 0 to 4.65 (H = .99, F = .01), with 0 meaning no sensitivity at all (chance level). Since d' is a distance measure, a d' score of 2 means twice as sensitive as a d' score of 1. Typical d' scores are up to 2, and a d' score of 1 roughly corresponds to 69% accuracy for both different and same trials.

(4)   d' = z(H) − z(F),

where both H and F represent performance ratios. The Hit Rate (H) is the ratio of "different" trials for which participants responded "different" when hearing HT3 + T4/T1 + T4 words or T1 + T4/HT3 + T4 words. The False Alarm Rate (F) is the ratio of "same" trials for which participants responded "different" when hearing two T1 + T4 words or two HT3 + T4 words. z refers to z-transform.

Linear mixed-effects analysis was conducted on participants' d' scores with Language (Huaiyuan, Naïve-Huaiyuan) as a fixed factor and random intercepts by participant as the random effects structure. For Language, native Huaiyuan listeners were treated as the baseline to which non-Huaiyuan listeners were compared. Results showed that native Huaiyuan listeners were almost twice as sensitive as non-Huayuan listeners when distinguishing Huaiyuan HT3 + T4 words from T1 + T4 words, and this performance difference was reflected by the significant d' difference (Huaiyuan, $M = 1.98$; non-Huaiyuan, $M = 1.01$; β = −.964, $SE = .222$, $t = -4.354$, $p < .001$), suggesting that native language experience improved Huaiyuan tone/word

discrimination (see **Figure 4**). Therefore, it is highly likely that both surface acoustic information and Huaiyuan phonological knowledge were used by native Huaiyuan listeners in these two perception experiments.
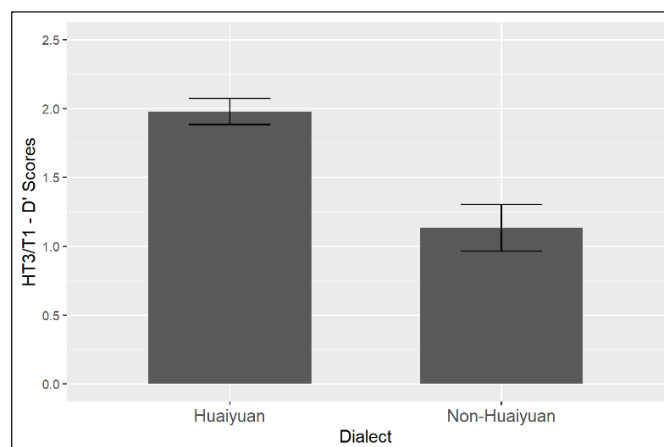

Figure 4: d' scores of the HT3 + T4 and T1 + T4 minimal pairs presented as boxplots. Error bars are plotted with a confidence level of .95.

4. Discussion

Huaiyuan Mandarin has three low-tone sandhis. T1 sandhi and T3 sandhi can be classified as the first type, which involves changing the low tone of the first syllable to a rising tone (T1 sandhi: ML + ML → MH + ML; T3 sandhi: MLM + MLM → MH + MLM), while half-third sandhi can be classified as the second type, which involves truncating the rising portion of the canonical T3 in the first syllable (MLM + MH/HL → ML + MH/HL). The current study investigated the production and perception of these low-tone sandhis in Huaiyuan Mandarin. Specifically, we tested whether Huaiyuan ST1 and T2, ST3 and T2, and HT3 and T1 were completely neutralized in production and perception, in order to examine the mechanisms underlying these low tone sandhis. Our acoustic and identification results showed comparable patterns for the ST1/T2 and HT3/T1 pairs, with the two tones of each pair not neutralized with each other. By contrast, the ST3/T2 pair showed different results in the two experiments, with nonneutralization in production, but complete neutralization in perception. Moreover, the discrimination results further revealed that native Huaiyuan listeners outperformed naïve Huaiyuan listeners in discriminating between Huaiyuan HT3 and T1, suggesting that native Huaiyuan listeners could use not only the acoustic details between the two tones of a minimal pair, but also native phonological knowledge to help identify and discriminate Huaiyuan words.

4.1. The mechanisms of Huaiyuan low-tone sandhis

For the first type, acoustic analysis of the ST1/T2 pair did not show neutralization, with ST1 having an average lower pitch, a steeper rising slope, and a more U-shaped contour than T2.

Similarly, the ST3/T2 pair did not display neutralization either, with ST3 showing an average lower pitch than T2. The results of these two tone sandhis are in line with those of the T3 sandhi in Standard Mandarin, in which case, the sandhi tone preserves some acoustic characteristics of its underlying tone (Wang & Li, 1967; Xu, 1997; Peng, 2000; Yuan & Chen, 2014; Tu & Chien, 2022). According to previous tone descriptions, Huaiyuan T1 is a low-falling tone and T3 is a low falling-rising tone, while T2 is a mid to high-rising tone (Gong, 2004; Shi, 2007; Zhang & Gong, 2010). The fact that both ST1 and ST3 have lower pitch height than T2 demonstrates that ST1 and ST3 keep their underlying low-tone feature. Additionally, in order for the low-falling T1 to reach its rising sandhi target, it has to rise drastically from the second half of the tone, inducing shape differences between ST1 and T2. These production patterns are consistent with those elicited from the T3 sandhi in Standard Mandarin, where ST3 has a lower pitch height and a later turning point than T2.

Interestingly, although native Huaiyuan listeners were unable to differentiate between ST3 and T2, demonstrating a mismatch between production and perception, just as some of the previous identification results obtained from the T3 sandhi in Standard Mandarin (Wang & Li, 1967; Peng, 2000), their identification performance of the ST1/T2 pair was significantly higher than chance. The difference in the identification results may be due to the fact that in Huaiyuan the acoustic difference between ST1 and T2 is larger than that between ST3 and T2. Thus, it is easier for native Huaiyuan listeners to perceive the difference and use it for word identification. The different results can also be due to the choice of stimuli. Using the stimuli that were closest to the average pitch tracks produced by all the participants, Tu and Chien (2022) observed perceptual nonneutralization between the ST3 and T2 in Standard Mandarin, consistent with the present identification results of the Huaiyuan ST1/T2 pair. The T1 sandhi and T3 sandhi in Huaiyuan also mirror the T1 sandhi and T3 sandhi in Tianjin Mandarin, which reveal nonneutralization between the sandhi tones and their corresponding canonical tones, leaving traces of the tonal characteristics of their underlying tones (Zhang & Liu, 2011; Li & Chen, 2016). These findings together showed that the first type of low tone sandhis in Mandarin dialects lead to noncategorical tone changes in production, whereas, regarding native speakers' perception, a sandhi rule may be categorical when the acoustic difference between the sandhi tone and its corresponding canonical tone is too trivial (e.g., ST3 and T2 in Huaiyuan and Standard Mandarin), or when the experimental paradigm is not sensitive enough to capture listeners' behaviors.

A question then rises as to why ST1 seemed to differ more from T2 than ST3 from T2. This question can be explained by the farther distance between Huaiyuan T1 and T2 than between Huaiyuan T3 and T2, hypothesizing that Huaiyuan T2 is also the target tone that Huaiyuan T1 sandhi attempts to reach, as evidenced by the fact that ST1 and T2 did not seem to differ drastically in acoustics, and participants' A' scores in identifying ST1 + T1 and T2 + T1 words were under .6. According to Zhao's acoustic results (Zhao, 2024), the canonical forms of Huaiyuan

T1 and T3 should be denoted as 311 and 112 in their respective tone values. The different pitch directions between Huaiyuan T1 (311) and T2 (34) may make it more challenging for the speakers to achieve the T2 target during the production process of T1 sandhi, with the result that ST1 and T2 differed not only in height, but also in shape. The shape difference between ST1 and T2 also made them easier to identify (Li, 2023). An alternative hypothesis for the mechanism of Huaiyuan T1 sandhi is that the target tone that Huaiyuan speakers intend to reach during the T1 sandhi process is ST1 rather than T2. Since ST1 and T2 are phonologically different, Huaiyuan T1 sandhi does not bring about neutralization, suggesting distinct underlying mechanisms between Huaiyuan T1 sandhi and T3 sandhi. This hypothesis is in line with Yuan and Chen (2014), which analyzed telephone speech of Standard Mandarin, and claimed that ST3 is phonologically different from T2 in Standard Mandarin, despite them being acoustically similar.

Regarding the half-third sandhi in Huaiyuan, it could potentially give rise to neutralization between HT3 and T1 (low-falling), leading to a different situation from the half-third sandhi in Standard Mandarin, which does not have a low-falling tone in its tonal inventory. Our acoustic results showed that Huaiyuan HT3 and T1 were not neutralized, with HT3 having a lower pitch, a smaller initial decline, and being more U-shaped and less S-shaped than T1, preserving the first half of the tonal contour of T3. Interestingly, according to the impressionistic tone descriptions, the half-third sandhi in Huaiyuan Mandarin bears an even stronger resemblance to the half-third sandhi in Tianjin Mandarin than to Standard Mandarin. In Tianjin, low-rising T3 is altered to low T1 when followed by T2 or T4, indicating that HT3 and T1 could also be neutralized. While Zhang and Liu (2011) characterized Tianjin half-third sandhi as nonneutralizing, their acoustic analysis revealed that the pitch track of Tianjin HT3 exhibited some degree of rising towards the end of a syllable, indicating incomplete truncation of a full rising tone in a syllable with insufficient duration. This suggests that Tianjin half-third sandhi may not be fully implemented during production. In contrast, the current half-third sandhi results of Huaiyuan revealed that the pitch track of HT3 consistently declined from the onset to the offset, without evidence of final pitch raising. Given that the pitch tracks of Tianjin T3 and Huaiyuan T3 appear remarkably similar, yet their respective HT3 realizations are different, (Zhang & Liu, 2011; Li & Chen, 2016; Zhao, 2024), we contend that the half-third sandhi processes in Huaiyuan and Tianjin are mechanically different, with Huaiyuan exhibiting more complete and consistent application of half-third sandhi. In perception, the identification and discrimination results revealed that native Huaiyuan listeners were very sensitive to the acoustic differences between HT3 and T1, obtaining an average A' score of .849 (i.e., middle-aged and younger groups combined) for identification and an average d' score of 1.98 for discrimination. These results showed that Huaiyuan half-third sandhi leads to noncategorical tone change, in line with the half-third sandhi in Standard Mandarin (Zhang & Lai, 2010).

As for why these Mandarin tone sandhis are nonneutralizing, we believe that it is due to them being phonologically transparent. For phonologically transparent sandhi rules, such as the T3 sandhi in Standard Mandarin, since native speakers are fully aware of the underlying tone from which the sandhi tone is derived (Chien et al., 2016; Chien et al., 2021; Nixon et al., 2015), during the process of sandhi production, the underlying tone influences the surface realization of the sandhi tone (e.g., Zhang & Lai, 2010). By contrast, for phonologically opaque sandhi rules (Kiparsky, 1973), such as the circular-chain-shift sandhi in Taiwanese, the surface sandhi form may be directly represented in the mental lexicon (Chien et al., 2017). Therefore, during production, the sandhi form is directly accessed and produced, leading to complete neutralization in pitch height and contour (Chien & Jongman, 2019; Myers & Tsay, 2008). Similar to the T3 sandhi in Standard Mandarin, Huaiyuan T1 sandhi and T3 sandhi are phonologically transparent, resulting in nonneutralization. These nonneutralizing sandhis in Huaiyuan imply that the canonical representation of Huaiyuan sandhi words may be stored in the mental lexicon. During speech production, native Huaiyuan speakers employ the computational mechanism, converting the underlying representation to the surface form, which requires more cognitive efforts (Zhang et al., 2015). Another representation view is also possible. That is, in addition to T1 and T3, their variant forms are also stored in the mental lexicon, which is similar to what was proposed by Yuan and Chen (2014) as well as Li and Chen (2015). Both studies claimed that the mental representation of T3 in Standard Mandarin should be T3 and T3 variant. Future studies should be conducted to tease apart these two proposals for Huaiyuan Mandarin.

In terms of whether the three low tone sandhis in Huaiyuan can be considered phonological rules, we argue that the T1 sandhi and T3 sandhi can be justified as phonological dissimilation, since both of them involve changing a low tone to a rising tone when followed by the same tone, and the sandhi tones are not predictable from their corresponding underlying tones (Li & Chen, 2016). Furthermore, several Mandarin studies have demonstrated that regressive coarticulation typically results in raising a preceding high tone by a low-onset tone (Shih, 1986; Shen, 1990; Xu, 1994, 1997; Zhang & Liu, 2011). The reverse pattern observed in Huaiyuan T1 sandhi and T3 sandhi, where word-initial low tones are converted to higher rising tones, suggests that these two Huaiyuan sandhis are less likely to stem from regressive coarticulation. With regard to Huaiyuan half-third sandhi, we propose that it may not need to be deemed a phonological rule since the sandhi process does not create a great mismatch between the underlying T3 and the surface HT3, with the surface HT3 preserving the first half of its base, T3.

4.2. Language experience on tone perception

In order to examine whether native Huaiyuan listeners could use not only the surface acoustic information, but also native phonological knowledge to perceive Huaiyuan tone sandhi words and their corresponding nonsandhi words, we compared native Huaiyuan and naïve Huaiyuan speakers' discrimination performances on the minimal pairs of HT3 + T4 and T1 + T4 words.

The results of native Huaiyuan listeners outperforming naïve Huaiyuan listeners showed that native language experience must have played a role in aiding native listeners in discriminating these two tones. The production results of the current study have illustrated that HT3 and T1 are acoustically different. In addition, the perception results revealed that although Huaiyuan HT3 and T1 are both low tones, they belong to two different toneme categories. Therefore, native Huaiyuan listeners could employ both phonetic and phonological knowledge to differentiate between HT3 and T1, obtaining relatively high d' scores.

The naïve Huaiyuan listeners are native listeners of a Mandarin dialect other than the dialects in the Jianghuai Mandarin family. Even though they do not speak Huaiyuan, they are still tone language speakers, therefore able to transfer their knowledge of tones in their native dialects to Huaiyuan tone discrimination, showing sensitivity in nonnative tone perception (Wayland & Guion, 2004; Chen et al., 2016; Choi & Tsui, 2022). However, their nonnative tone knowledge was not enough to allow them to be on a par with native Huaiyuan listeners. Furthermore, according to the perceptual assimilation model (PAM; Best, 1995; PAM-S; So & Best, 2010, 2014), if two nonnative phonemes assimilate into one native phoneme category with equal goodness of fit, discriminability would be low. Since the Mandarin dialects of the naïve Huaiyuan listeners have only one low tone, the HT3 and T1 in Huaiyuan would assimilate into the single low-tone category of their native Mandarin dialects, making it more difficult for the naïve Huaiyuan listeners to differentiate between the two tones. These factors together may have contributed to the average d' score of 1.03, approximately 70% accuracy, in the discrimination experiment, which was decent, but still significantly poorer than that acquired by the native listeners.

5. Conclusion

In conclusion, the present study first utilized a production experiment to show that Huaiyuan low tone sandhis lead to noncategorical tone changes. By examining the acoustic findings across different Mandarin dialects, we propose that Huaiyuan T1 sandhi and T3 sandhi should be classified as phonological, while Huaiyuan half-third sandhi is more difficult to justify as phonological, and can instead be explained by a phonetic mechanism. Second, the identification results further demonstrated that only Huaiyuan T3 sandhi is perceptually neutralizing, but T1 sandhi and half-third sandhi are perceptually nonneutralizing. Together, the production and perception findings suggest distinct underlying mechanisms among these low tone sandhis. Lastly, we also demonstrated that native Huaiyuan listeners outperformed naïve Huaiyuan listeners in discriminating between Huaiyuan HT3 and T1, suggesting the recruitment of phonological knowledge in native tone perception. Future studies can investigate the production and perception of low tone sandhis in the other Mandarin dialects, so that the mechanisms of low tone sandhis can be better understood not only from phonetic and phonological viewpoints, but also from a typological viewpoint.

Appendix A. Critical stimuli

ST1 + T1	天机	荒灾	摊开	沧州	拼单	荒山	巫山	青天	秋千	猫砂
Translation	a hidden sign of fate	famine	unfold	City Cang	joint order	barren mountain	Mountain Wu	blue sky	swing	Cat litter
T2 + T1	田鸡	蝗灾	弹开	常州	凭单	黄山	吴山	晴天	求签	毛纱
Translation	Frog	locust disaster	bounce out of	City Cang	voucher	Mountain Huang	Mountain Wu	clear sky	pray and draw divination sticks at a temple	yarn
IPA	tʰian tɕi	xuɑŋ tsɛ	tʰan kʰɛ	tsʰɑŋ tsəu	pʰin tan	xuɑŋ san	u san	tɕʰin tʰian	tɕʰiəu tɕʰian	mɔ sa

ST3 + T3	讨喜	碾米	彩纸	五里	彩礼	五组	起码	忍者	反锁	土改
Translation	pleasure	rice milling	colored paper	five kilometers	betrothal gifts	five sets	at least	ninja	be locked in	land reform
T2 + T3	淘洗	黏米	裁纸	无礼	财礼	无阻	骑马	仁者	繁琐	涂改
Translation	wash	sticky rice	cut paper	rude	monetary price	unimpeded	ride a horse	benevolent people	cumbersome	erase
IPA	tʰɔ ɕi	nian mi	tsʰɛ tsʅ	u li	tsʰɛ li	u tsu	tɕʰi ma	zən tsə	fan suɔ	tʰu kɛ

HT3+T4	企盼	紧闭	小气	止步	总部	死路	款待	缴付	审视	影像
Translation	hope for	shut tightly	stingy	halt	headquarters	dead end	hospitality	pay	Examine	image
T1+T4	期盼	金币	消气	支部	中部	丝路	宽带	交付	绅士	音像
Translation	expect	gold coin	cool down	branch	central section	Silk Road	broad-band	deliver	gentleman	audio visual
IPA	tɕʰi pʰan	tɕin pi	ɕiɔ tɕʰi	tʂɿ pu	tsoŋ pu	si lu	kʰuan tɛ	tɕiɔ fu	sen ʂɿ	in tɕʰiaŋ

Funding

Competing interests

The authors have no competing interests to declare.

Authors' contributions

JZ: writing original draft, data collection, and data analysis. HY: conceptualization, finalizing the manuscript, and revisions. Y-FC: conceptualization, data analysis, writing and finalizing the manuscript, supervision, and revisions. All authors approved the submitted version.

References

Bates, D., Maechler, B., Bolker, B., & Walker, S. (2015). Fitting linear mixed effects models using lme4. *Journal of Statistical Software, 67*, 1–48.

Best, C. T. (1995). *A direct realist view of cross-language speech perception (W. Strange, Ed.)*. York Press.

Boersma, P., & Weenink, D. (2023). Praat: Doing Phonetics by Computer [Computer program]. http://www.praat.org/

Chao, Y. R. (1930). A system of tone letters. *La Maître Phonétique, 45*, 24–27.

Chao, Y. R. (1968). *A Grammar of Spoken Chinese*, Chapter 1, 25–30. University of California Press, Berkeley.

Chen, A., Liu, L., & Kager, R. (2016). Role of native language in cross domain pitch perception. *Language, Cognition, and Neuroscience, 31*, 751–760. https://doi.org/10.1080/23273798.2016.1156715

Chen, M. Y. (2000). *Tone sandhi: Patterns across Chinese dialects*. Cambridge University Press.

Chen, Z. (1996). Putonghua cihui guifan wenti [The issues of lexical regulation in Putonghua]. *Studies of the Chinese Language, 3*, 194–205.

Chien, Y.-F., & Jongman, A. (2019). Tonal neutralization of Taiwanese checked and smooth syllables: An acoustic study. *Language and Speech, 62*(3), 452–474. https://doi.org/10.1177/0023830918785663

Chien, Y.-F., Sereno, J. A., & Zhang, J. (2016). Priming the representation of Mandarin tone 3 sandhi words. *Language, Cognition, and Neuroscience, 31*, 179–189. https://doi.org/10.1007/s10936-020-09745-0

Chien, Y.-F., Sereno, J., & Zhang, J. (2017). What's in a word: Observing the contribution of both underlying and surface representations. *Language and Speech, 60*(4), 643–657. https://doi.org/10.1177/0023830917690419

Chien, Y.-F., Yan, H., & Sereno, J. A. (2021). Investigating the lexical representation of Mandarin Tone 3 phonological alternations. *Journal of Psycholinguistic Research, 50*(4), 777–796. https://doi.org/10.1007/s10936-020-09745-0

Choi, W., & Tsui, K. Y. (2022). Perceptual integrality of foreign segmental and tonal information: Dimensional transfer hypothesis. *Studies in Second Language Acquisition, 45*(4), 1–18. https://doi.org/10.1017/S027226312200051

Dmitrieva, O., Jongman, A., & Sereno, J. A. (2010). Phonological neutralization by native and non-native speakers: The case of Russian final devoicing. *Journal of Phonetics, 38*(3), 483–492. https://doi.org/10.1016/j.wocn.2010.05.001

Francis, A. L., & Ciocca, V. (2003). Stimulus presentation order and the perception of lexical tones in Cantonese. *Journal of the Acoustical Society of America, 114*(3), 1611–1621. https://doi.org/10.1121/1.1603231

Geng, J. (2007). Anhui Huaiyuan fangyan yinxi [Phonology of Huaiyuan dialect of Anhui]. *Journal of Fuyang Normal University (Social Science Edition), 5*, 79–82.

Gong, G. (2004). *Anhui Huaiyuan fangyin diaocha baogao* [Report on a Survey of HuaiYuan Dialect in AnHui Province] [Master's thesis]. Ningxia University, Ningxia. https://www.cnki.net/kns/defaultresult/index

Grier, J. B. (1971). Nonparametric indexes for sensitivity & bias: computing formulas. *Psychological Bulletin, 75*(6), 424–429.

Herd, W., Jongman, A., & Sereno, J. A. (2010). An acoustic and perceptual analysis of /t/ and /d/ flaps in American English. *Journal of Phonetics, 38*(4), 504–516. https://doi.org/10.1016/j.wocn.2010.04.003

Kiparsky, P. (1973). Abstractness, opacity and global rules. In O. Fujimura (Ed.), *Three dimensions of linguistic theory* (pp. 57–86). Tokyo, Japan: Tokyo Institute for Advanced Studies of Language.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Li, Q. (2023). A preliminary study on the online processing of anticipatory tonal coarticulation – Evidence from eye movements. *Frontiers in Psychology, 14*, 1137095. https://doi.org/10.3389/fpsyg.2023.1137095

Li, Q., & Chen, Y. (2016). An acoustic study of contextual tonal variation in Tianjin Mandarin. *Journal of Phonetics, 54*, 123–150.

Li, X. (2004). Hanyu fangyan liandubiandiao de cengji he leixing [Typology of tone sandhi in Chinese]. *Fangyan (Dialect), 1*, 16–33.

Li, X., & Chen, Y. (2015). Representation and processing of lexical tone and tonal variants: evidence from the mismatch negativity. *PLoS One, 10,* e0143097. https://doi.org/10.1371/journal.pone.0143097

Lin, Y.-H. (2007). *The sounds of Chinese.* Cambridge University Press.

Liu, J. (1994). Jinan fangyan shangshangxianglian qianzi biandiao de shiyanfenxi [Experimental analysis of the first syllable of disyllabic T3 sandhi words in Jinan dialect]. *Studies in Language and Linguistics, 2*, 76–84.

Liu, L. (2004). *Hanyushengdiaolun* [On Chinese tones]. *Journal of School of Chinese Language and Culture Nanjing Normal University, 4*, 189.

Luo, J. (2016). Putonghua shi woguo minzuronghe de chanwu-qianxi Hanyu Putonghua de laiyuan [Mandarin Chinese is a product of our nation's ethnic integration – A brief analysis of the origins of Standard Chinese]. *Modern Chinese, 12*, 114–116.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge University Press.

Mei, T.-L. (1977). Tones and tone sandhi in 16th century Mandarin. *Journal of Chinese Linguistics, 5*, 237–260.

Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R*. Taylor & Francis.

Myers, J., & Tsay, J. (2008). Neutralization in Taiwan Southern Min tone sandhi. *Interfaces in Chinese Phonology, 49*(1), 47–78.

Nixon, J. S., Chen, Y., & Schiller, N. O. (2015). Multi-level processing of phonetic variants in speech production and visual word processing: evidence from Mandarin lexical tones. *Language, Cognition, and Neuroscience, 30*, 491–505. https://doi.org/10.1080/23273798.2014.942326

Norman, J. (1988). *Chinese*. Cambridge University Press.

Peirce, J. W., Hirst, R. J., & MacAskill, M. R. (2022). *Building Experiments in PsychoPy (2nd ed.)*. Sage.

Peng, S.-H. (2000). Lexical versus 'phonological' representations of Mandarin sandhi tones. In M. B. Broe & J. B. Pierrehumbert (Eds.), *Papers in laboratory phonology V: Acquisition and the lexicon* (pp. 152–167). Cambridge University Press.

Shen, X. (1990). Tonal coarticulation in Mandarin. *Journal of Phonetics, 18*, 281–295.

Shen, X. (1992). Hanyu Putonghua shengdiao de xietongfayin [Tonal coarticulation in Mandarin Chinese]. *Contemporary Linguistics, 2*, 26–32.

Shi, S. (2007). *Jianghuaiguanhua rushing yanjiu* [Research on the checked tones in Jianghuan Mandarin] [PhD dissertation]. Beijing Language and Culture University, Beijing. https://www.cnki.net/KCMS/detail/detail.aspx?dbcode=CDFD&dbname=CDFD9908&filename=2007162154.nh&uniplatform=OVERSEA&v=VX4LGDk3wnBEqLKhpQnwHZZllbc_gBneeLt34qeTrN5X7v8RyF-eOQORF9a1eQOm

Shih, C.-L. (1986). *The prosodic domain of tone sandhi in Chinese* [PhD dissertation]. University of California, San Diego. https://www.proquest.com/openview/160fd2cc458e962bb3a2e7aa909426a0/1?pq-origsite=gscholar&cbl=18750&diss=y

Shih, C.-L. (1997). Mandarin third tone sandhi and prosodic structure. In J. Wang & N. Smith (Eds.), *Studies in Chinese phonology* (pp. 81–123). Mouton de Gruyter.

Snodgrass, J. G., Levy-Berger, G., & Haydon, M. (1985). *Human Experimental Psychology*. Oxford University Press.

So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech, 53*, 273–293. https://doi.org/10.1177/0023830909357156

So, C. K., & Best, C. T. (2014). Phonetic influences on English and French listeners' assimilation of Mandarin tones to native prosodic categories. *Studies in Second Language Acquisition, 36*, 195–221.

Tagliaferri, B. (2019). *Paradigm*. http://www.paradigmexperiments.com/index.html

Tu, J.-Y., & Chien, Y.-F. (2022). The role of categorical perception and acoustic details in the processing of Mandarin tonal alternations in contexts: An eye-tracking study. *Frontiers in Psychology, 12*, Article 756921. https://doi.org/10.3389/fpsyg.2021.756921

Wang, W. S.-Y., & Li, K.-P. (1967). Tone 3 in Pekinese. *Journal of Speech and Hearing Research, 10*, 629–636.

Wang, Y., Yu, M., & Wu, Q. (2018). Zhongyin he yunlu dui Beijinghua lianshangbiandiao de zuoyong [The effects of stress and prosody on the T3 sandhi in Beijing]. *Chinese Journal of Phonetics, 2*, 72–82.

Warner, N., Jongman, A., Sereno, J. A., & Kemper, S. (2004). Various factors underlying the perception of acoustic cues to word boundaries in English. *Journal of Phonetics, 32*(2), 173–196. https://doi.org/10.1016/S0095-4470(03)00032-5

Wayland, R. P., & Guion, S. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning, 54*(4), 681–712. https://doi.org/10.1111/j.1467-9922.2004.00283.x

Wu, Y. (2020). Guanhua fangyan shuangyinjie lianshangbiandiao de leixing ji chengyin [Typology and formation of disyllabic T3 sandhi words in Mandarin dialects]. *Linguistic Sciences, 19*(2), 207–218.

Wu, Z. (2002). Zhongguo yinyunxue he yuyinxue zai yanyu yanyuhechengzhong de yingyong [Application of Chinese phonology and phonetics in Chinese speech synthesis]. *Language Teaching and Linguistic Studies, 1*, 1–14.

Xu, Y. (1994). Production and perception of coarticulated tones. *Journal of the Acoustical Society of America, 95*, 2240–2253.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics, 25*, 61–83.

Xu, Y. (2005/2010). ProsodyPro.praat. Available online at: http://crdo.fr/crdo000723

Yuan, J., & Chen, Y. (2014). 3rd tone sandhi in standard Chinese: A corpus approach. *Journal of Chinese Linguistics, 42*(1), 218–237.

Zee, E. (1980). A spectrographic investigation of Mandarin tone sandhi. *UCLA Working Papers in Phonetics, 49*(9), 98–116.

Zhang, A., & Gong, G. (2010). Anhui Huaiyuan fangyan de rusheng yanbian ji guishu [Historical variation and categorization of the checked tones in Huaiyuan dialect of Anhui Province]. *Journal of Hebei University (Philosophy and Social Science), 35*(4), 1–9.

Zhang, C., Xia, Q., & Peng, G. (2015). Mandarin third tone sandhi requires more effortful phonological encoding in speech production: Evidence from an ERP study. *Journal of Neurolinguistics, 33*, 149–162.

Zhang, J., & Lai, Y. (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology, 27*, 153–201. https://doi.org/10.1017/S0952675710000060

Zhang, J., & Liu, J. (2011). Tone sandhi and tonal coarticulation in Tianjin Chinese. *Phonetica, 68*, 161–191.

Zhang, W. (2014). Xiandai Hanyu Putonghua zhuang, chuang, shuang de laiyuan, xingchengguocheng ji xiangguan wenti [The source, formation, and related questions of zhuang, chuang, shuang in Modern Chinese Putonghua]. *Journal of Nanyang Normal University, 13*(11), 36–39.

Zhao, J. (2024). *Huaiyuanhua zhonghebiandiao shiyanyanjiu* [An Acoustic and Perceptual Study of Huaiyuan Dialect Sandhi] [Master's thesis]. Fudan University, Shanghai.

Zhu, X. (2010). *Phonetics*. The Commercial Press.