

The Acquisition of Mandarin by Heritage Speakers and Second Language Learners: The Case of Tone Sandhi Perception*

Chung-yu CHEN

**Master's & Doctor's Program in Teaching Chinese as a Second Language
National Chengchi University**

Chilin SHIH

**Department of East Asian Languages and Cultures,
Department of Linguistics
University of Illinois at Urbana-Champaign**

Abstract

This study investigates how heritage speakers (HSs) and late learners of a second language (L2ers) of Mandarin (both English-dominant) perceive Mandarin tone 3 (T3) sandhi. T3 sandhi occurs when two T3 syllables occur consecutively, and the first syllable becomes a T2. When a T3 syllable precedes a non-T3 syllable, the T3 syllable becomes a half-T3. Forty-two Mandarin native speakers (NSs), 21 HSs and 25 L2ers completed a Tone Identification Task. Participants listened to two monosyllabic sound files separately and were then asked to choose the disyllabic sound file (one out of four) that corresponded to NS production of the disyllabic sequences. While the

* Acknowledgement: This article is based on Chapter 3 of the first author's dissertation at the University of Illinois at Urbana-Champaign (UIUC), supervised by Dr. Tania Ionin and supported by National Science Foundation (NSF) (BCS-1844447) and UIUC SLATE Doctoral Research Award. Results have also been reported at the 2021 International Annual Conference of Teaching Chinese as a Second Language (in Taipei). We thank the members of the UIUC experimental linguistics group, the dissertation committee members, and the reviewers for helpful comments and suggestions. All remaining errors are ours.

differences are not always statistically significant, the trend is that (1) HSs performed better than L2ers overall (HSs being more native-like than L2ers); (2) while HSs have more difficulty identifying T3 sandhi and half-T3 conditions than other conditions, L2ers seems to show overall difficulty with disyllabic tones.

Keywords: English-dominant heritage speakers, half tone 3, L1-English L2 learners, perception, tone 3 sandhi

1. Introduction

Research has shown that the Mandarin tonal system is difficult for second language (L2) learners who acquire Mandarin later in life, no matter whether their first language (L1) is a tonal language like Cantonese or a non-tonal language like English (e.g., Hao 2012). Yet little research has been done on those who acquired the Mandarin tonal system very early on, but later switched their language dominance to a different language. These speakers, who are commonly called “heritage speakers” (HSs), are defined as individuals who were exposed to a home language other than the societal language (i.e., English in the U.S.). They were exposed to the home language from their parents from birth, but as adults, they are dominant in the societal language, with varying degrees of proficiency in their home/heritage languages.

Broadly, researchers of heritage languages are interested in knowing whether HSs pattern more like monolingual native speakers (NSs) or late learners of an L2 (L2ers), and what causes these differences or similarities (e.g., age of acquisition, quality, and quantity of the exposure). For example, the well-known Critical Period Hypothesis (see Mayberry and Kluender 2018, for an overview) has argued that age of acquisition (AoA) contributes to the different outcomes of L1 and L2 acquisition. The Critical Period Hypothesis in L2 acquisition predicts that pre-puberty learners can achieve native-like proficiency in an L2 while post-puberty learners cannot, due to decreasing brain plasticity from biological maturation. Though the existence of a critical period in L2

acquisition is debatable, researchers agree that age effects exist and affect phonetics and phonology more than morphosyntax (e.g., Granena and Long 2013). If so, given sufficient input, HSs should acquire the sound system like monolingually-raised NSs and outperform adult L2ers because they acquired their home language early in life, despite a later reduction in input/exposure (because another societal language co-exists).

By now, most studies have found HSs to outperform L2ers on speech perception, but not necessarily on production (e.g., Korean: Oh et al. 2003; Mandarin: Chang et al. 2011; Chang and Yao 2016; Spanish: Au et al. 2002, 2008; Knightly et al. 2003; Kim 2020). For example, Kim (2020) found that Spanish HSs have an advantage over L2ers in perception, but not production, possibly because they hear Spanish more than they speak it. Early exposure to a language gives learners an advantage in perceiving sounds even when they have no conscious recollection of it. For example, using functional MRI, Pierce et al. (2014) found that, unlike French monolinguals, international children adopted by French-speaking Canadian parents and who had stopped hearing Chinese by age one or two still showed brain responses to tones similar to those produced by Chinese-French bilinguals.

Tone 3 (T3) sandhi is chosen in this article for several reasons: (a) while there are many studies on T3 sandhi, the acquisition of T3 sandhi, particularly by HSs and L2ers, remains understudied; (b) it is a later-acquired phenomenon in monolingual children compared with basic tonal categories, so we can see whether the HS advantage reported in other languages will hold; and (c) it is a complex phenomenon which is typically not emphasized in the classroom, unlike basic tones.

While L2-Mandarin tones have long received considerable attention (e.g., see Zhang 2018a, for a review), recent studies have begun to examine T3 sandhi by L2ers (e.g., Chen et al. 2019). However, few if any have examined the perception of Mandarin tones and tone sandhi by HSs of Mandarin. The present study investigates the perception of Mandarin tones and tone sandhi by HSs and L2ers of Mandarin, focusing on disyllabic sequences.

We structure the paper as follows. Section 2 introduces the properties of

Mandarin tones and tone sandhi and the acquisition literature. Next, we ask the research questions in Section 3. Section 4 presents the methodology, including participants and materials. Results and discussion are presented in Sections 5 and 6, respectively. Finally, Section 7 concludes the article.

2. Mandarin Tone Sandhi

2.1 Properties of Mandarin Tone Sandhi

Mandarin has four lexically contrastive tones: high-level (T1), high-rising (T2), low-falling-rising (or low-falling; T3), and high-falling (T4). Additionally, there is a neutral tone on an unstressed syllable, with a shorter duration and a pitch determined by the preceding tones; we will not focus on this tone in this article, besides briefly mentioning that L2ers sometimes misperceived another specific tone (to be discussed later) as a neutral tone. Using the five-point pitch values developed by Chao (1930), T1 is transcribed as [55], T2 as [35], T3 as [214] (two other variants – [21] and [35] – are discussed below), and T4 as [51]. As a standard notation, the pitch values are in brackets. In *Pinyin* (Romanized script), the tone numbers (1, 2, 3, and 4) are indicated after the syllables, e.g., *ma1* stands for *ma* pronounced with T1. The syllable *ma* with T1 though T4 are thus *ma1*, *ma2*, *ma3* and *ma4*, which mean *mother*, *hemp*, *horse* and *scold* respectively in Mandarin. An alternate way is to indicate the tone on the vowel, as in *mā*, *má*, *mǎ*, and *mà*. Table 1 summarizes the lexical tones in Mandarin.

Table 1: Mandarin Lexical Tones

	Height/Contour	Pitch Value	Examples
T1	high-level	[55]	<i>ma1</i> ‘mother’
T2	high-rising	[35]	<i>ma2</i> ‘hemp’
T3	low-falling-rising or low-falling	[214] or [21]; [35] under T3 sandhi	<i>ma3</i> ‘horse’
T4	high-falling	[51]	<i>ma4</i> ‘scold’

Tone sandhi is the tonal alternation in natural speech. The T3 sandhi rule is that, when two underlying T3 syllables occur consecutively, the first syllable becomes T2 [35] (see e.g., Shih 1986 for more analyses). For example, the

Mandarin greeting ‘*ni3 hao3*’ (‘you good’) is pronounced as ‘*ni2 hao3*’ (but *ni3* and *hao3* if pronounced separately). In addition to disyllabic sequences, tone sandhi also applies in multi-syllabic sequences, introducing more complexity because the prosodic and the morpho-syntactic-domains interact (e.g., Speer et al. 1989; Tu et al. 2017). When an underlying T3 syllable precedes a non-T3 syllable, namely T1, T2, or T4, the T3 syllable is pronounced as a half-T3 ([21]; a low-falling tone). While T3 sandhi is phonological in nature and has no straightforward phonetic explanation, the half-T3 rule is phonetically motivated in that T3 is changed from [214] to [21], without the final rise to reduce the articulatory effort (see e.g., Zhang and Lai 2010). Unlike the T3 sandhi rule, where T3 is changed to a categorically different tone (i.e., T2), the half-T3 is not categorically different from T3; see (1). In this article, we use “full-T3” [214], “sandhied T3” [35] (the T3 that undergoes tone sandhi and is realized as T2) and “half-T3” [21] to describe the three phonetic variants/realizations of T3, while using “T3” to mean the citation/underlying form represented in speakers’ minds.

(1) Mandarin T3 sandhi rules

Full-T3 sandhi rule: T3 [214] → T2 [25] / __ + T3

Half-T3 (sandhi) rule: T3 [214] → low-falling tone [21] / __ + T1/2/4

While many NSs and instructed learners are aware of the full T3 sandhi rule, few are aware of the half-T3 rule, despite half-T3 having the widest distribution. While full-T3 [214] is traditionally described as occurring in isolation and in utterance-final positions, Beijing Mandarin uses half-T3 in utterance-final positions (Duanmu 2000), and Taiwanese Mandarin even allows half-T3 in isolation (Tai 1978). Currently, it is debated whether the underlying form of T3 is indeed the full-T3 or if it is actually the more-widely distributed half-T3. The debate has implications for theoretical phonology and pedagogy (e.g., how T3 should be taught, which variant should be taught first, etc.) (see e.g., Zhang 2017). We will refer to T3 as the underlying form represented in speakers’ minds since it is not the goal of the present article to solve this debate.

While some perceptual studies have shown that NSs cannot reliably differentiate between a T2 and a sandhied T3 (e.g., Peng 2000), some studies

have found the opposite (e.g., Lin and Hsu 2018; Tu and Chien 2020). In production, some acoustic studies have found differences between a T2 and a sandhied T3 (e.g., Yuan and Chen 2014; Li et al. 2021). For example, testing adult NSs on both perception and production using pseudo-words and real words that are minimal pairs of similar frequency (*er2yu3* ‘baby talk’ vs. *er3yu3* ‘whisper’), Zhang and Peng (2013) found no (significant) differences in perception of either pseudo-words or real words; in contrast, acoustic differences were found from pseudo-words only. Comparing sandhied T3 and half-T3, Zhang and Lai (2010) found that NSs made fewer mistakes in half-T3 than sandhied T3, because the former is more phonetically motivated.

2.2 L1 Acquisition of Mandarin Tone Sandhi

In L1 acquisition, lexical tone perception is developed before age one (see Tsao 2016, for a review). Children acquire T1, T4, T2, and finally T3. Generally, tone errors are rare beyond age two, though some have reported children who did not master all four tones by 2;6 (for reviews, see Zhu 2016; Tsay 2016). However, Wong and Strange (2017) recently found that tone production by children as old as six was still rated significantly lower in accuracy compared to adults. Children in Wong and Strange (2017)’s study also made more tone errors in first syllables in a word than in second syllables.

While earlier studies found that T3 sandhi is acquired by age three based on perceptual data transcribed by NSs (e.g., Zhu 2002; Huang 2006), recent acoustic studies have found that this is not the case (Tang et al. 2019; Xu Rattanasone et al. 2018). For example, Tang et al. (2019) found that, while 3-year-olds could productively apply the T3 sandhi rule to novel disyllabic words, children up to even the age of 5 differed from adults when applying this rule to trisyllabic words.

2.3 L2 Acquisition of Mandarin Disyllabic Tones and Tone Sandhi

For L2ers, tones are notoriously difficult (for reviews, see Wang et al. 2006, 2016; Zhang 2018a). Given that perception difficulty and production difficulty of tones do not always correlate, after briefly discussing both production and perception, we focus on tone perception by English speakers in the next two

paragraphs, as the present study examines the perception of HSs and L2ers who are English-dominant. Generally, T1 and T4 are acquired earlier, and produced and perceived with higher accuracy than T2 and T3 (for reviews, see e.g., Zhang 2018a). While some studies have found that learners who speak a tonal L1 such as Cantonese outperform those who speak a non-tonal L1 such as English in Mandarin perception (e.g., Lee et al. 1996), Cantonese NSs in Hao (2012) did not outperform English NSs in neither perception nor production of Mandarin tones.

In tone perception, monosyllabic T4 is the easiest to identify in isolation (or in utterance-final position) for L1-English L2ers because its falling pitch is acoustically similar to the end of English declaratives (e.g., Wang et al. 2006). Due to acoustic similarity (similar F0 contours) between T2 and T3, these two tones are difficult to discriminate for both tonal and non-tonal L2ers, and under some circumstances, even for NSs (e.g., Chang 2011). For example, Hao (2012) found that both Cantonese and English NSs have problems differentiating T2 and T3 in Mandarin, with Cantonese NSs having additional difficulty distinguishing Mandarin T1 and T4.

Recent studies have found that L1-English L2ers performed better on monosyllabic tones than disyllabic tones (e.g., He and Wayland 2013), even at the advanced level (e.g., Pelzl et al. 2019). For disyllabic tones (e.g., Hao 2018; Pelzl et al. 2019), T4 becomes the hardest for L1-English L2ers to identify when presented in non-final position due to interference from English intonation. The mutual confusion between T2 and T3 persists for disyllabic sequences in both first and second syllables (e.g., Chen 1997, cited in Hao 2018). Tone in first syllables is also more difficult to identify than in second syllables (e.g., Hao 2012), similar to findings from Mandarin-speaking children (Wong and Strange 2017) mentioned above. Yet for the inexperienced learners (i.e., beginner, just 3 months of study) in He and Wayland (2013), while T2 and T3 are significantly easier to identify as the second syllable than as the first syllable, T1 is significantly easier to identify as the first syllable than as the second syllable; there are no significant differences on T4. For the relatively “experienced learners” (i.e., 12 months of study) in He and Wayland (2013), it is only T3 that is significantly easier to identify as the second syllable than as the first syllable,

while there are no significant differences on T1, T2 and T4. He and Wayland (2013:16) attributed both inexperienced and experienced learners' difficulty with first syllable T3 as due to "low-falling" tone, which we call half-T3 in the present study.

While some L2 studies on disyllabic sequences exclude the T3T3 combination from the analysis due to T3 sandhi (e.g., He and Wayland 2013; Hao 2018; Yang and Yang 2019), several studies have begun to examine T3 sandhi in L2ers, mostly in production: Table 2 summarizes the L2 studies on T3 sandhi that have English-speaking learners, including both production and perception studies (for a study with Korean-speaking L2ers, see Li et al. 2021; for a study with Cantonese-speaking L2ers, see Tu et al. 2017). While correct production was possible when evaluated by NSs (e.g., Yang 2016), Chen et al.'s (2019) acoustic study found that L2ers' production is not native-like. To our knowledge, the only L2-perception study on T3 sandhi is Zhang (2017, see also 2018b). With L1-English L2ers, both Yang (2016) and Jin (2019) found higher accuracy in production for half-T3 than full-T3 (significant differences in Yang's study, but not in Jin's) while Zhang (2017; see also Zhang 2018a, 2018b) found the opposite. Thus, this issue is not settled yet. While the T3 sandhi rule is more commonly taught than the half-T3 rule (Yang and Jin 2018; Jin 2019), Jin (2019) found that L2ers performed better on half-T3 than sandhied T3 and attributed this finding to the stronger phonetic motivation of the half-T3 than sandhied T3. Specifically, Zhang (2018b) found that half-T3 (low-falling) [21] is mostly misperceived as T4 (high-falling) [51] by beginner and intermediate level learners, but not advanced level learners because both have falling contours; furthermore, half-T3 is also sometimes misperceived as the neutral tone because both half-T3 and the neutral tone have a short duration. Similarly, in production, half-T3 (low-falling) [21] is mispronounced as T4 (high-falling) [51] because both have falling contours. In He and Wayland (2013:19), T3 as the first syllable before T1, T2, and T4 was only correctly identified by the experienced learners as a T3 at 43.4% and mis-identified as T1 at 8.8%, T2 at 16.2% and T4 at 31.6%, similar to Zhang's finding. In contrast, when T3 was the second syllable after T1,

T2, and T4, the experienced learners correctly identified it as a T3 at 86.5% and mis-identified it as a T2 at 13.5%.

Table 2: Selected Studies on Mandarin T3 Sandhi that Included L1-English L2ers

	L1s	Task Format	Pattern Summary
Yang (2016; Ch 6)	English	Production	Significantly higher accuracy of half-T3 than T3 sandhi
Zhang (2018b; Ch 6)	L1-English/ Japanese/Korean L2ers on production; L1-English L2ers on production and perception	Production and perception	Higher accuracy for T3 sandhi than half-T3 in production; L2ers were unable to transcribe half-T3 as T3 in perception
Chen et al. (2019)	Cantonese & English	Production	L2ers' production is not native-like, with the Cantonese group outperforming the English group
Jin (2019)	English	Production	Higher accuracy for half-T3 than T3 sandhi, but no significant differences; similar accuracy for half-T3 in T3T4, T3T2, and T3T1

2.4 Heritage Language Acquisition of Mandarin Tones

While heritage phonology is a burgeoning field (see Chang 2021, for an overview, e.g., Cantonese: Tse 2016; Korean: Chang and Weiss-Cowie 2021; Spanish: Kim 2019), studies on heritage Mandarin remain relatively rare. Some research comes from Chang and colleagues and Yang (2015). Chang and colleagues found that HSs have an advantage over L2ers in producing Mandarin vowels, plosives, retroflex (Chang et al. 2011) and tones (Chang and Yao 2016), but not with neutral tones (Chang and Yao 2019).¹ Examining trisyllabic

¹ Chang and Yao (2019) found that L2ers patterned more like NSs than HSs in some (but not all) of the measures on neutral tones. However, as acknowledged by the authors, since neutral tones in non-obligatory contexts vary across dialects (more consistent in northern Mandarin but not in southern Mandarin), L2ers' advantage over HSs in neutral tones may be due to (1) L2ers' familiarity with the standard northern Mandarin commonly taught in L2 classrooms, and (2) some HSs were HSs of southern Mandarin.

sequences, Yang (2015) also found that HSs outperformed L2ers in recognizing the starting point of tones in their perceptual space. A related study by Tsukada et al. (2015) found that HSs of Cantonese did not outperform L1-English L2ers in discriminating monosyllabic Mandarin tones and even had additional problems differentiating between Mandarin T1 and T4, which they mapped to Cantonese T1, similar to findings from L1-Cantonese L2-Mandarin learners (Hao 2012). To our knowledge, no study has examined the perception of T3 sandhi by HSs or compared HSs and L2ers for this phenomenon.

2.5 Pedagogical Note

Given that all L2ers and most HSs in the present study were at some point instructed learners, a note on the pedagogy is in order. Tones are usually introduced in the first class, along with Romanization and tone marks. According to a recent survey (Yang and Jin 2018), 86% of instructors taught the T3 sandhi rule, while only 53% of them taught the half-T3 rule. Even when these rules are taught, tone errors are not necessarily corrected in class. While most instructors adopt the ‘full-T3 first’ teaching method by (initially) teaching T3 as a full T3 (likely due to the theoretical assumption that full-T3 is the base form of T3), some advocate the ‘half-T3 first’ teaching method by just teaching T3 as a low tone for at least two reasons: first, to help beginning learners differentiate T2 and a half-T3, since L2ers were more likely to be confused between T2 and full-T3 (both T2 and full-T3 have rising tones); secondly, to eliminate the need to introduce the half-T3 sandhi rule (see the reviews in Zhang 2017, 2018b). Examining the L2 production and perception data, Zhang (2017, 2018b) argues that the dominant ‘full-T3 first’ teaching method led some L2ers to over-produce a full T3 when it should be a half-T3 or when it should be a sandhied T3 (i.e., T2). In the latter case, because both T2 and full-T3 have rising tones, sometimes NSs may misjudge full-T3 produced by L2ers as T2, inflating the accuracy rate of T3 sandhi application. While not specifically discussing sandhied T3, He et al. (2016) compared these two teaching methods and tested students’ production on monosyllabic words, disyllabic words and sentences which contained T3 (production accuracy rated by judges based on perception). They found that the

students receiving the full-T3 teaching method (termed “pitch direction-focused” in their article) performed better on monosyllabic words, but students receiving the half-T3 method (termed “pitch height-focused” in their article) performed better on sentences. The two groups patterned similarly on disyllabic words (including half-T3 contexts). Given that tones rarely occur in isolation in conversation and that students receiving the half-T3 method were better at generalizing their production accuracy from learned words to new words, He et al. (2016) recommend the pitch height-focused teaching method and call for different teaching methods to be tested on sandhied T3 contexts.

3. Research Questions and Hypotheses

The current study asks two research questions (RQs):

RQ 1: Can HSs and L2ers whose dominant language is English correctly identify sandhied T3 and half-T3 sandhi in Mandarin tone perception?
Is one easier than the other?

RQ 2: How do HSs and L2ers differ in their identification of sandhied T3 and half-T3 sandhi in Mandarin tone perception?

To answer these questions, we used a Tone Identification Task (with stimuli recorded by a NS of Mandarin) to test whether HSs and L2ers know the full T3 sandhi rule and the half-T3 sandhi rule by correctly identifying T2 as sandhied T3 and half-T3 as T3. Given prior studies, we hypothesize that NSs would perform at ceiling while HSs and L2ers might not. Comparing the two variants, we hypothesize half-T3 sandhi to be easier to identify than sandhied T3 due to stronger phonetic motivation on half-T3 sandhi than sandhied T3, even though the former is less commonly taught. Alternatively, if sandhied T3 is easier to acquire than half-T3 sandhi, it may be due to explicit instructional effort on T3 sandhi rule. Given that Mandarin T3 sandhi is acquired before age three in monolingual children, HSs are expected to have acquired it fully because most HSs are still regularly exposed to Mandarin at home by that age. For L2ers, even though most of the instructed L2ers are taught the full T3 sandhi rule, they may

still have difficulty acquiring it since phonology is known to be difficult for adult L2ers. Thus, a HS advantage is expected.

4. Method

4.1 Procedure

All tasks were administered on Qualtrics, a web-based survey tool. Participants completed, in order, a background questionnaire (for HSs and L2ers, in English; for NSs, in Chinese), a Chinese proficiency test, a Tone Production Task (not reported here), a Tone Perception Task, and several other tasks (not reported here).² The participants' linguistic backgrounds were collected using the questionnaire. While the proficiency test aimed to independently measure participants' Mandarin proficiency in order to match the proficiency of heritage and L2 groups, the test required participants to read Chinese characters, and was a better measurement for grammar than overall proficiency. This is a limitation of the present study since we did not control for proficiency using an aural or oral task (e.g., an oral picture-naming task in Kim 2020). All tasks were untimed. On average, NSs spent about 12 minutes on the Tone Identification Task, while HSs and L2ers spent about 15 minutes.

Within each trial, participants listened to two separate sound files, each with one monosyllable. The two sound files were placed on one page, with the first monosyllable above the second. After participants listened to the two sound files of monosyllables, they clicked "Next" to move to the next page with four sound files of disyllabic sequences, made up of the two monosyllables but with different tonal combinations. The four sound files were placed on the page horizontally (in random order except for the critical condition, discussed below), from left to right. Participants needed to choose the disyllabic sequence (one out of four) that corresponded to the way that a NS would pronounce the previous two individual monosyllables together (i.e.,

² The production task (not reported here) preceded the perception task to avoid biasing participants' production and raising their awareness of the T3 sandhi phenomena. It is possible that some participants became aware of the T3 sandhi phenomena in the tone production task, or later in the tone perception task.

participants were asked to “[c]hoose one which you believe is the native pronunciation” instead of the “correct” pronunciation). They were told that none of the disyllabic sequences in any tonal combinations were real words, though most (if not all) individual monosyllables were real words.

There were two practice trials (*yuan2la1*, *zai4ni3* – these two tonal combinations were not tested). Given that participants were not able to return to the previous page to listen to the monosyllables again, in the practice trials, the first author (who tested the participants) emphasized that they had to “remember the sounds [the monosyllables they listened to]” before clicking “Next” and until they had chosen the disyllabic sound files in that trial.³ The first author indicated that they could just repeat the monosyllables in their heads without saying them out loud, but some still repeated them quietly to themselves. They were not allowed to take notes on the monosyllables, because they might have written down tone notations such as “33” or “✓✓” for the T3T3 condition (the critical condition - more explanation below), which would have likely encouraged them to choose the unnatural-sounding T3T3 sound files.

Participants were allowed to listen to the sound files as many times as they wanted, but whether they actually listened to all of the sound files was not controlled. Some NSs chose the correct disyllabic file after identifying the correct one; some did not necessarily start with the leftmost sound files. Most HSs and L2ers did listen to all four sound files (even multiple times) before making a decision.

4.2 Participants

Data from 42 Mandarin NSs, 21 HSs, and 25 L2ers were included, after excluding seven participants based on the language background questionnaire. HSs who listed a non-Mandarin Chinese language as one of their native languages (two with Cantonese, one with Taiwanese, and one with

³ As one reviewer points out, this design does not take into consideration the individual difference on short-term memory. Participants could have forgotten the tones or even the entire syllables and had to randomly choose one out of the four sound files to finish that trial.

Shanghainese) were excluded because extensive experiences with other tonal languages since childhood may differentiate them from other Mandarin HSs.

In terms of testing locations, NSs were tested in Beijing ($n = 22$) and Taipei ($n = 20$) in a private setting. All NSs were born and raised in China or Taiwan, had not spent more than one year abroad, had not been immersed in a bilingual environment such as an English-speaking international school, and were not students of linguistics or Chinese pedagogy. All HSs and L2ers reported in the present study completed the tasks in the United States (HSs: $n = 15$; L2ers: $n = 10$) or Taiwan (HSs: $n = 6$; L2ers: $n = 15$) in a private setting. HSs tested in Taiwan grew up in the United States or Canada but moved to Taiwan as adults or were visiting Taiwan at the time of testing. The participants' relevant background information is summarized in Table 3.

Table 3: Information about the Participants

	NSs ($n = 42$)	HSs ($n = 21$)	L2ers ($n = 25$)
Age of testing	Mean 22.3 (range 19 - 37)	Mean 21.5 (range 19 - 33)	Mean 27 (range 20 - 46)
Age of Acquisition of Mandarin (in years)	N/A	Mean: 0.05 (range 0-1) 20 since birth 1 at 1	Mean: 19.3 (range 8-31) 3 at ages 8 through 12 4 ages 14 through 17 18 ages 18 and up
Age of Acquisition of English (in years)	N/A	Mean: 1.6 (range 0-5) 13 since birth 8 before or at age 5	N/A
Average years of Mandarin classes	N/A	Mean: 6.2 (range 0.7-14)	Mean: 3.8 (range 0.4-13)
Age of arrival in the United States	N/A	16 born in the United States 1 at age 2 1 at age 4 3 born in Canada	19 born in the United States, but 1 grew up in the United Kingdom 6 born outside the United States ^a

Note: ^a1 born in Canada; 1 born in Bahamas, but grew up in Canada; 4 born in the United Kingdom.

4.3 Materials

Seven conditions were created; see Table 4. Conditions refer to the underlying/citation tones. While T3T3 must be realized as T2T3 due to tone sandhi when pronounced together, the condition is named T3T3, referring to T3T3 disyllabic sequence in underlying tones, even though the correct answer would be T2T3. For example, after listening to *da3* and *yi3* (*da3* being placed above *yi3*, but on the same page), the answer for *da3yi3* is *da2yi3* (out of the four possible choices: *da1yi3*, *da2yi3*, *da3yi3*, and *da4yi3*) due to T3 sandhi. T3T1, T3T2, and T3T4 conditions are necessary to establish that the T3 sandhi rule only applies before another T3 and not before T1, T2, or T4. Similarly, T1T3 and T4T3 conditions are necessary to establish that the T3 sandhi rule only applies when T3 (as a first syllable) is before another T3 and does not apply to T1 and T4 before a T3. For the T2T3 condition, while T3 sandhi does not apply, the outcome is the same.

Table 4: Conditions (Correct Answers **Bolded**)

Underlying Tone	Sound Files	Purposes of the Conditions
T3T3	T1T3, T2T3 , T3T3, T4T3	T3 sandhi
T2T3	T1T3, T2T3 , T3T3, T4T3	This condition has T3 as the second syllable; while the T3 sandhi rule does not apply, the realization is the same as the T3T3 condition
T1T3	T1T3 , T2T3, T3T3, T4T3	These two conditions have T3 as the second syllable to demonstrate that the T3 sandhi rule only applies to T3 (as a first syllable) and not to T1 and T4 before T3
T4T3	T1T3, T2T3, T3T3, T4T3	
T3T1	T1T1, T2T1, T3T1 , T4T1	These three conditions have T3 as the first syllables to demonstrate that the full T3 sandhi rule only applies before another T3 and not before T1, T2, or T4
T3T2	T1T2, T2T2, T3T2 , T4T2	
T3T4	T1T4, T2T4, T3T4 , T4T4	

With five tokens per condition, there were 35 trials, each presented with four sound files. Five common Mandarin syllables (*yi*, *ma*, *da*, *you*, and *ke*) that have

four different tones (not every Mandarin syllable can be pronounced in all four tones) were chosen. Vowels (such as the close front rounded vowel [y]) and consonants (such as the voiceless alveolo-palatal affricate [tɕ]) that are known to be difficult for English speakers were avoided. Before imposing tones, there were five disyllabic sequences: *yima*, *dayou*, *yike*, *keda*, and *dayi*. Then, each of them was imposed with the seven tonal combinations. For example, *dayi* was tested as *da1yi3*, *da2yi3*, *da3yi3*, *da4yi3*, *da3yi1*, *da3yi2*, and *da3yi4*. None of the disyllabic sequences were real words.

There were five blocks, each with seven conditions. Blocks were randomized, but the trials within blocks were not (as participants who saw the T3T3 and T2T3 conditions consecutively might have become aware of the purpose of the task). The four sound files for each trial were presented in random order except for the T3T3 condition. For the T3T3 condition, to avoid participants from just choosing T3T3 without having the chance to hear T2T3, the sound files of the unnatural-sounding T3T3 were placed before (i.e., on the left of) the sound files of T2T3 in two tokens (in two different blocks) and after the sound files of T2T3 in two other tokens. The fifth token had a random order. For the full list of blocks and stimuli, see the Appendix.

In each of the seven conditions, there were four types of sound files. Note that the second syllable of all four types of sound files in each condition were the same tone. (Ideally, there should be $(4*4=)16$ choices, but the present design reduced the choices participants had to make.) In four (out of the seven) conditions (T3T3, T1T3, T2T3, and T4T3; the top four rows in Table 4), the second syllable was always T3 in all sound files. For the other three conditions (T3T1, T3T2, and T3T4; the bottom three rows in Table 4), the second syllable for the corresponding conditions are T1, T2, and T4 respectively. Thus, participants can answer correctly as long as they can identify which tone the first syllable has, without identifying the second syllable, because all four sound files in a given trial have the same second syllable.

4.4 A Note on the Sound Materials (Recording)

The stimuli were recorded by a female NS of (Taiwanese) Mandarin (the

second author) in a sound-attenuated booth. While T3 in Taiwanese Mandarin is typically realized as half-T3 [21], even in isolation, she deliberately produced clear, full-T3 [214] in monosyllables, which is easier for listeners to identify. While it is generally very difficult for a NS to produce T3T3 sequences without any pause in between, the speaker was able to produce it without a pause. The first author later listened to the recordings and checked the pitch contour in Praat to confirm that the first syllable in all T3T3 sequences was indeed T3 (technically, a half-T3), not T2. The stimuli were not synthesized nor modified.⁴

4.5 Predictions

For the critical T3T3 condition, recall that T3 becomes T2 if followed by another T3. The prediction is that, when asked to choose a disyllabic sequence after hearing two T3 syllables, those who do not know T3 sandhi will choose the sound file of T3T3, which sounds unnatural to NSs. For HSs and L2ers, the T2T3 condition is expected to be slightly easier than T3T3, since they do not need to know the T3 sandhi rule. We also predict greater accuracy on the half-T3 (i.e., T3T1, T3T2, T3T4) conditions over the T3T3 condition because half-T3 is phonetically motivated (Yang 2016; Jin 2019; but see Zhang 2018b).

Next, we discuss what it means when participants correctly choose T2T3 for the T3T3 condition but may do so due to overgeneralization seen from mistakes in other conditions. Note that the second row through fourth row in Table 5 all have T3 as the second syllable (i.e., T1T3, T2T3, and T4T3), and the last three rows all have T3 as the first syllable (i.e., T3T1, T3T2, and T3T4). For the conditions ending with T3 syllables, if participants still chose T2T3 sound files for T1T3 and T4T3 conditions, this would indicate overgeneralization of T3

⁴ One reviewer asks whether it would be better if the stimuli were not produced entirely naturally. To have more controlled stimuli, we did once consider synthesizing two isolated syllables and perhaps modifying the duration of all disyllabic sequences to be comparable to a typical disyllabic sequence. Yet, if all stimuli were made this way, we were concerned if such unnaturalness would negatively affect participants' performance. Additionally, it looked like Zhang (2018b: 90) also used natural-occurring stimuli in her perception task on T3 sandhi. We do not currently have a definite answer whether using synthesized and/or modified stimuli would yield results different from the present study; we will leave this possibility for future research.

sandhi to other disyllabic sequences that have T3 as the second syllable or misperception of T1/T4 as T2. For the conditions starting with a T3 (i.e., T3T1, T3T2, and T3T4), if participants correctly choose T2T3 for the T3T3 condition, but wrongly choose T2T1, T2T2, and T2T4 respectively for the T3T1, T3T2, and T3T4 conditions, this would indicate overgeneralization of T3 sandhi to other disyllabic sequences that have T3 as the first syllable or misperception of T3 as T2.

For the three conditions ending with T3 (T1T3, T2T3, and T4T3) vs. the three conditions starting with T3 (T3T1, T3T2, and T3T4), the latter is predicted to be more difficult. Greater accuracy on the T1T3, T2T3 and T4T3 conditions over the half-T3 conditions is expected due to half-T3 rules (see He and Wayland 2013). First, T2 and T3 are mutually confusable, so HSs and L2ers might choose T2T1, T2T2, and T2T4 respectively for the T3T1, T3T2, and T3T4 conditions. Second, given that T3 before T1, T2, and T4 is pronounced as half-T3 ([21]; low falling), it may be perceived as T4 ([51]; high falling), as found in Zhang (2017) by low-level (but not advanced-level) learners. If so, HSs and L2ers might incorrectly choose T4T1, T4T2, and T4T4 for the T3T1, T3T2, and T3T4 conditions. These predictions are summarized in Table 5.

Table 5: Predictions for the Tone Identification Task (Correct Answers **Bolded**)

Underlying Tone	Sound Files	NSs	HSs and L2ers
T3T3	T1T3, T2T3 , T3T3, T4T3	T2T3	T3T3 if they do not know the full T3 sandhi rule
T1T3	T1T3 , T2T3, T3T3, T4T3	T1T3	T2T3 if they overgeneralize the T3-to-T2 rule to any disyllabic sequence ending with a T3 syllable or if they misidentify T1 as T2
T2T3	T1T3, T2T3 , T3T3, T4T3	T2T3	T2T3
T4T3	T1T3, T2T3, T3T3, T4T3	T4T3	T2T3 if they overgeneralize the T3-to-T2 rule to any disyllabic sequence ending with a T3 syllable or if they misidentify T4 as T2

Table 5: Predictions for the Tone Identification Task (Correct Answers **Bolded**) (cont.)

Underlying Tone	Sound Files	NSs	HSs and L2ers
T3T1	T1T1, T2T1, T3T1 , T4T1	T3T1	T2T1 if they overgeneralize the T3-to-T2 rule to any disyllabic sequence starting with a T3 syllable or simply misperceive T3 as T2; T4T1 if they misidentify half-T3 as T4
T3T2	T1T2, T2T2, T3T2 , T4T2	T3T2	T2T2 if they overgeneralize the T3-to-T2 rule to any disyllabic sequence starting with a T3 syllable or simply misperceive T3 as T2; T4T2 if they misidentify half-T3 as T4
T3T4	T1T4, T2T4, T3T4 , T4T4	T3T4	T2T4 if they overgeneralize the T3-to-T2 rule to any disyllabic sequence starting with a T3 syllable or simply misperceive T3 as T2; T4T4 if they misidentify half-T3 as T4

For HS/L2 comparisons, HSs are predicted to outperform L2ers in all conditions, including the T3T3 condition (T3 sandhi). Given that monolingual Mandarin-speaking children acquire T3 sandhi by age three, HSs presumably also acquire it even when living in an English-speaking country since their Mandarin exposure should still be extensive before the age of three.

5. Results

5.1 Group Analysis

A correct response was coded as “1” and an incorrect response as “0”. (Note that T2T3, but not T3T3, was coded as the correct answer for T3T3 condition here; choices of T2T3 vs. T3T3 will be reported later.) Then, the raw scores in each condition (range 0-5 as there were five tokens per condition) were averaged across the participants and converted to percentages. Figure 1 shows the mean

accuracy of the group results.⁵

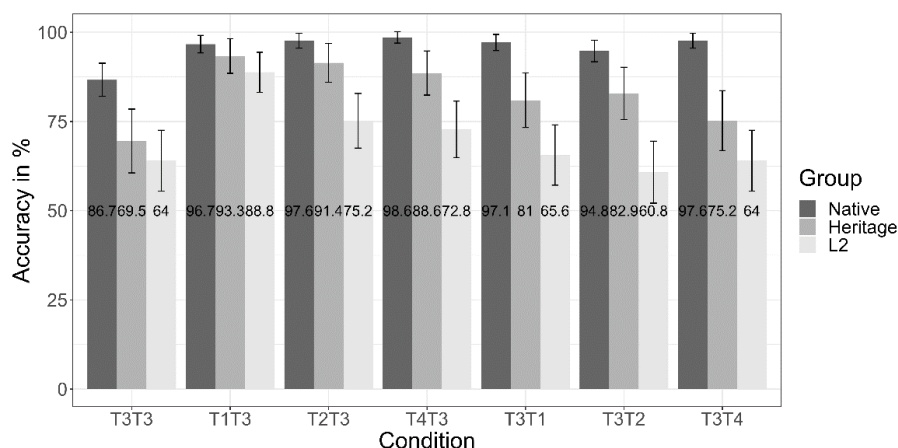


Figure 1: Group Results: Mean Accuracy (in %)

Data were analyzed in a logistic mixed-effects model (Jaeger 2008) using the *glmer()* function in the *lme4* package in R (R Core Team 2019). The model included group (NSs, HSs, and L2ers) and condition (T3T3, T1T3, T2T3, T4T3, T3T1, T3T2, and T3T4) and their interaction as fixed effects; the random effects included a random intercept for subjects and a random intercept for items. Dummy coding was used. The reference level for group was the NSs, while the reference level for condition was the T3T3 condition. We then used the *Anova()* function in the *car* package to assess the overall effect. There is a main effect of

⁵ As one reviewer suggests, dialect may be a factor. Given that we did not have detailed information on which dialect HSs and L2ers learn (e.g., dialects the instructors speak), we coded only NSs' dialects based on testing location and report the figures below:

	T3T3	T1T3	T2T3	T4T3	T3T1	T3T2	T3T4
Speakers of Mainland Mandarin (<i>n</i> = 22)	91.8%	99.1%	99.1%	99.1%	99.1%	97.3%	100%
Speakers of Taiwanese Mandarin (<i>n</i> = 20)	81%	94%	96%	98%	95%	92%	95%

We initially tried to run a logistic mixed-effect model on only NSs (with dialect and condition as fixed effects; subject and item as random effects), but encountered convergence issues. After removing both random effects, we ran a logistic model on only NSs: condition ($\chi^2(6) = 20.23, p < .01$) and dialect ($\chi^2(1) = 5.38, p = .02$) are significant, but not the interaction. It appears that speakers of Taiwanese Mandarin scored lower overall (which may be due to not paying enough attention or the setting of the testing location being more distracting etc.). Since both dialect groups showed similar pattern (lowest accuracy on the T3T3 condition and almost at-ceiling performance on the other), we did not further discuss dialect effect.

group ($\chi^2 (2) = 14.86, p < .001$), with NSs being more accurate than HSs, who in turn, were more accurate than L2ers. There is also a main effect of condition ($\chi^2 (6) = 47.52, p < .001$), with overall lower accuracy in the T3T3 condition and higher accuracy in the T1T3 condition. The two-way interaction of condition and group is significant ($\chi^2 (12) = 34.37, p < .001$), with HSs and particularly L2ers being more accurate on T1T3 condition than other conditions.

Pairwise comparisons were conducted via *emmeans* (Lenth et al. 2019) following the significant interactions; the *p*-values are significant at the Tukey-adjusted alpha level of .05. We will discuss the between-group differences before the within-group differences. The only condition in which all three groups patterned similarly was the T1T3 condition (NSs vs. HSs: $p = .78$; NSs vs. L2ers: $p = .28$; HSs vs. L2ers: $p = 1$). For the T3T3 condition, HSs patterned with NSs ($p = .21$) and L2ers ($p = 1$); NSs were much more accurate than L2ers (87% vs. 64%), although the difference is marginal ($z = 3.51, p = .06$). NSs were more accurate than HSs in four (out of seven) conditions: the differences were significant on T4T3 ($z = 3.99, p < .05$), T3T1 ($z = 4.51, p < .01$), and T3T4 ($z = 5.22, p < .0001$), although only marginal on T3T2 ($z = 3.48, p = .07$). In addition to T1T3, HSs patterned similarly with NSs on T2T3 ($p = .18$). NSs were more accurate than L2ers in six (out of seven) conditions: the differences were significant on T2T3 ($z = 5.05, p < .001$), T4T3 ($z = 5.41, p < .0001$), T3T1 ($z = 5.83, p < .0001$), T3T2 ($z = 5.71, p < .0001$), and T3T4 ($z = 6.05, p < .0001$), though only marginal on T3T3 ($z = 3.51, p = .06$). HSs patterned with L2ers in all conditions (all $p > .5$).

For within-group differences (the baseline is the T3T3 condition), NSs were significantly more accurate on all conditions than on the T3T3 condition (T3T3 vs. T1T3: $z = -3.98$; T3T3 vs. T2T3: $z = -4.30$; T3T3 vs. T4T3: $z = -4.48$; T3T3 vs. T3T1: $z = -4.15$; T3T3 vs. T3T2: $z = -3.22$; T3T3 vs. T3T4: $z = -4.30$; all $p < .05$), with the exception being T3T2. For this T3T2 condition, NSs had similar accuracy rates as for the T3T3 condition ($p = .14$). HSs' T3T1, T3T2 and T3T4 conditions showed similar accuracy rates as the T3T3 condition (all $p > .5$). However, compared with the T3T3 condition, HSs were significantly more accurate on the T1T3 and T2T3 conditions (T3T3 vs. T1T3: $z = -4.33$; T3T3 vs.

T2T3: $z = -4.02$, both $p < .01$), and marginally more accurate on T4T3 conditions ($z = -3.50$, $p = .06$). L2ers were more accurate on the T1T3 condition than the T3T3 condition ($z = -4.90$, $p < .001$); the T3T3 condition had scores similar to the other five conditions (all $p > .5$).

5.2 Error Analyses for the T3T3, T3T1, T3T2, and T3T4 Conditions

Because NS did not uniformly choose T2T3 in the T3T3 condition, further analysis was conducted on their responses to examine whether they instead chose T3T3 in the T3T3 condition. Additionally, it is informative to see whether HSs and L2ers also chose T3T3 in the same condition or if they chose the other two entirely incorrect options. Seven error analyses, each for one condition, were conducted, but only the T3T2, T3T1, T3T2, and T3T4 conditions were discussed further, since specific error patterns were predicted. Furthermore, error analysis conducted with the three conditions ending with T3 (i.e., T1T3, T2T3, and T4T3) revealed that the errors were evenly distributed among the three incorrect choices and/or participants had low error rates.⁶ Due to lack of obvious patterns, we do not discuss them further.

Figure 2 and Figure 3 each presents a percentage breakdown of the answer choices for the T3T3 and the T3T1 condition by group. In Figure 2, NSs chose the T3T3 sound file 11% of the time while HSs and L2ers did so over 25% of the time. L2ers additionally chose both T1T3 and T4T4 about 5% of the time.⁷ In

⁶ For the T1T3 condition, 3.8% of HS answers were incorrect on T2T3, 1% on T3T3, 1.9% on T4T3; 5.6% of L2er answers were incorrect on T2T3, 0.8% on T3T3, and 4.8% on T4T3. For the T2T3 condition, 5.7% of HS answers were incorrect on T1T3, 1.9% on T3T3, and 1% on T4T3; 10.4% of L2er answers were incorrect on T1T3, 8.8% on T3T3, and 1.4% on T4T3. For the T4T3 condition, HSs incorrectly chose T1T3 1.9% of the time, T2T3 7.6% of the time, and T3T3 1.9% of the time; L2ers incorrectly chose T1T3 11.2% of the time, T2T3 8% of the time, and T3T3 8% of the time.

⁷ As one reviewer inquired, we report the statistical results on the error analysis for the T3T3 condition. All three groups performed similarly in their choices of T1T3 and T4T3 sound files (all $p > .5$); NSs differed significantly from HSs and from L2ers in choosing T2T3 (NSs vs. HSs: $z = 3.57$, $p = .02$; NSs vs. L2ers: $z = 4.71$, $p < .001$) and T3T3 sound files (NSs vs. HSs: $z = -3.52$, $p = .02$; NSs vs. L2ers: $z = -3.44$, $p = .03$), while HSs and L2ers did not differ significantly from each other (both $p > .5$). As one reviewer suggested, NSs must be using some cues that HSs and L2ers don't have to avoid this error.

Figure 3, as predicted, due to overgeneralization and/or mutual confusion between T3 and T2, HSs and L2ers each incorrectly chose T2T1 about 20% of the time. Additionally, about 10% of the answers made by L2ers mistook half-T3 as T4, thus resulting in the incorrect choosing of T4T1. HSs appear to have correctly identified half-T3 as T3 over 80% of the time. Similar patterns were found with T3T2 (Figure 4) and T3T4 (Figure 5): both HSs and L2ers incorrectly chose T2 when it was T3 about 15-25% of the time, with L2ers having higher error rates. In addition, only L2ers incorrectly identified half-T3 as T4 in disyllabic sequences starting with T3 about 7-10% of the time.

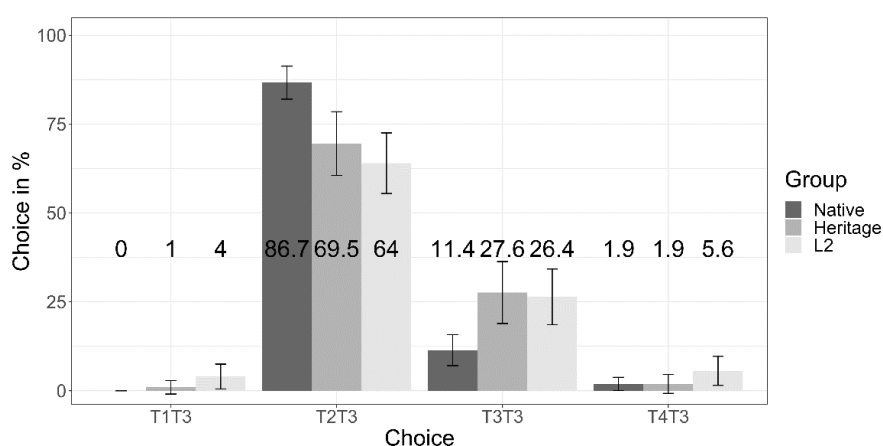


Figure 2: Error Analysis for the T3T3 Condition (Mean Choice (in %) by Type of Choice, Totaling 100% Over All 4 Types of Choice for Each Group)

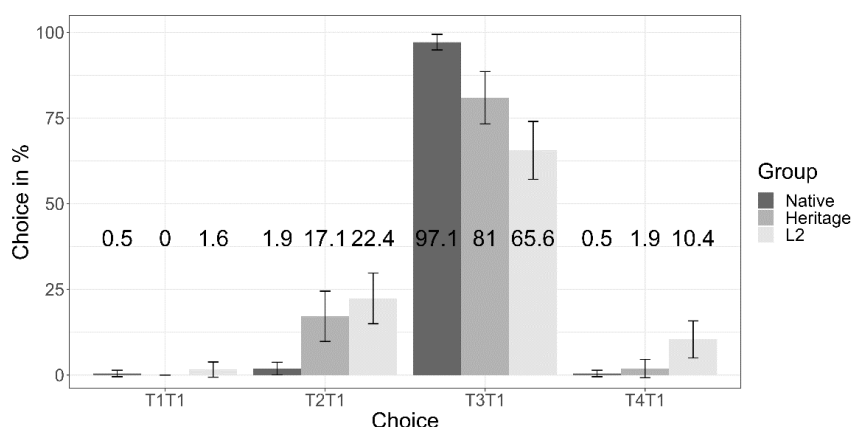


Figure 3: Error Analysis for the T3T1 Condition (Mean Choice (in %) by Type of Choice, Totaling 100% Over All 4 Types of Choice for Each Group)

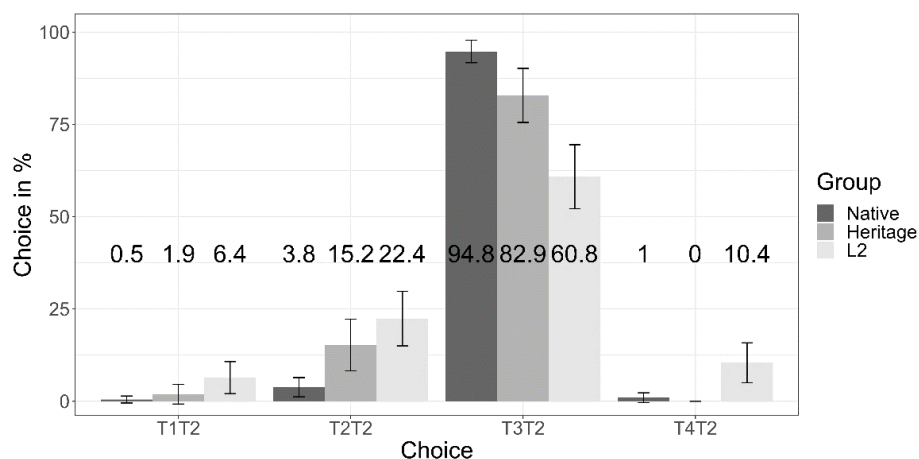


Figure 4: Error Analysis for the T3T2 Condition (Mean Choice (in %) by Type of Choice, Totaling 100% Over All 4 Types of Choice for Each Group)

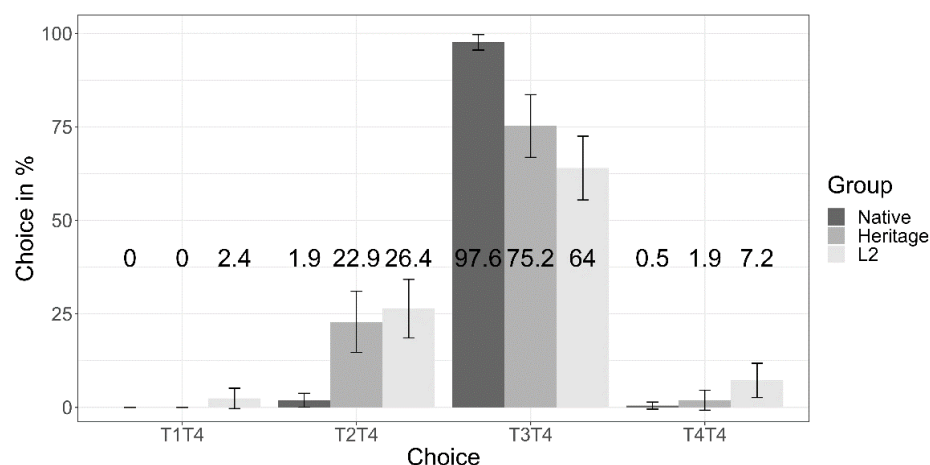


Figure 5: Error Analysis on the T3T4 Condition (Mean Choice (in %) by Type of Choice, Totaling 100% Over All 4 Types of Choice for Each Group)

5.3 Participants' Choices for the T3T3 Condition

To examine what tones participants chose for the T3T3 condition, we coded their preference into three categories: T2T3-preferred, T3T3-preferred, and others. If a participant chose T2T3 or T3T3 more than three times (≥ 3) out of five, they were classified as T2T3-preferred or T3T3-preferred respectively. The rest were classified as others. Figure 6 below shows the preferred choices for the T3T3 condition by each group. Eighty-eight percent of NSs, 71% of HSs and

64% of L2ers correctly chose T2T3. However, 10% of NSs, 24% of HSs, and 16% of L2ers (incorrectly) chose T3T3. Additionally, 20% of L2ers incorrectly chose one of the other two sound files - either T1T3 or T4T3. This is consistent with the group results in Figure 1 and Figure 2 where L2ers showed the lowest accuracy (when we coded only T2T3, but not T3T3, to be “accurate”).

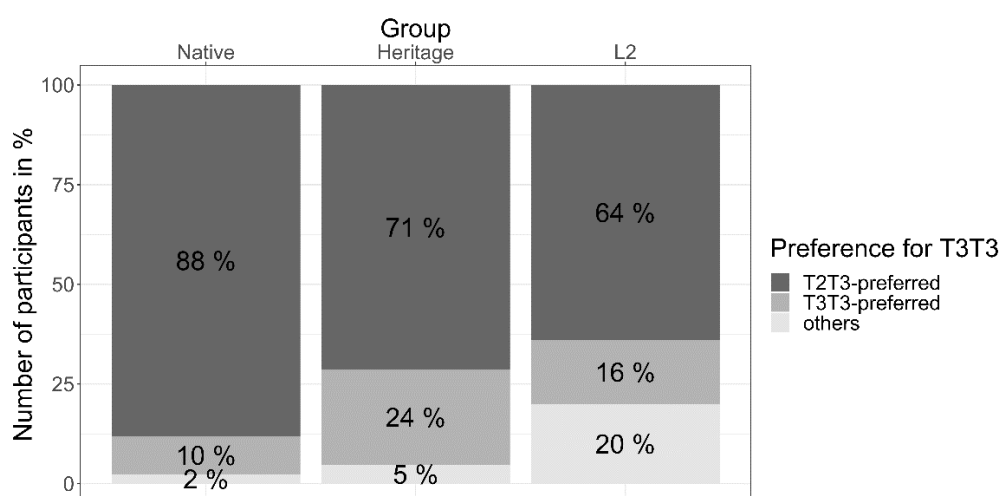


Figure 6: Participants' Choices on the T3T3 Condition by Group

6. Discussion

The unexpected results from NSs on the T3T3 condition are discussed before moving on to the performances by HSs and L2ers. After discussing the T3T3 condition, we discuss the other six conditions by HSs and L2ers (since NSs performed at ceiling). Next, we discuss the HSs/L2ers comparisons, and end with pedagogical implication.

6.1 T3T3 Condition by NSs, HSs, and L2ers

In Figure 6, while most (88%) NSs indeed chose T2T3 for at least three out of five tokens, 10% chose T3T3. It could be that these 10% of the NSs indeed think T3T3 is how it should be pronounced, but this is largely contradictory to what we know from previous studies and from the production task we conducted (not reported in the present study). Given that a T3 before another T3 cannot be naturally pronounced as a T3 (as it occurs in the current task), the fact that 10%

of the NSs in this study did not apply the T3 sandhi rule has to be due to some undesired task effects.⁸

One possible source of the task effect is the over-attention to the first syllables. As mentioned above, participants could choose the correct answer by only paying attention to the first syllable. If they did so, it may explain why they chose T3T3 for the T3T3 condition. The experiment was set up in a way to minimize the possibility of choosing T3T3. By having two separate pages, participants had to memorize the two monosyllables. If they pronounced the two monosyllables together (out loud or silently in their minds), the sequence should be T2T3. Admittedly, some participants might have memorized the tone notation number (i.e., T1/T2/T3/T4), which would likely encourage them to choose the T3T3 sound file.

Given that about 10% of the NSs choose T3T3 for the T3T3 condition, some might question whether it is indeed “incorrect”. However, there is no evidence for knowing T3 sandhi if T3T3 is counted as correct. While HSs patterned with NSs in the T3T3 condition, L2ers did not. For HSs and L2ers, the former chose T2T3 70% of the time and the latter chose T2T3 64% of the time. Using a three-token cutoff (out of five), 71% of HSs and 64% of L2ers choose T2T3 instead of T3T3. Thus, we could say that the majority of the HSs and over half of the L2ers did show clear evidence of knowing T3 sandhi. Data from the tone production task will provide complementary evidence for this.

6.2 Other Conditions by HSs and L2ers

Recall that both groups patterned with NSs in the T1T3 condition, but HSs additionally patterned with NSs in the T2T3 condition (as well as the T3T3 condition) while L2ers did not. Thus, while HSs patterned with L2ers in all conditions, HSs were more native-like than L2ers were.

For the two sandhi rules, HSs and L2ers do not perform differently on T3 sandhi and half-T3 sandhi conditions. The three conditions starting with T3

⁸ Another possibility is that individual differences on short-term memory resulted in the non-at-ceiling performance on the T3T3 condition, though participants scored at ceiling in other conditions.

(T3T1, T3T2, and T3T4) are predicted to be more difficult than the conditions ending with T3 (i.e., T1T3, T2T3, and T4T3) due to difficulty differentiating between T2 and T3/half-T3. This prediction is supported with HSs but less so with L2ers, who struggled with all conditions except for the T1T3 condition. L2ers' difficulty with disyllabic tones is consistent with previous studies (e.g., Hao 2012, 2018; Pelzl et al. 2019). Additionally, first syllables were more difficult than second or final syllables, which also happen with monolingual-speaking children (e.g., Wong and Strange 2017). Among the three conditions ending with T3, HSs and particularly L2ers were more accurate on the T1T3 condition than the T2T3 condition. Accuracy rates in these two conditions were in turn slightly higher than in the T4T3 condition. We suspect that the lower accuracy of the T4T3 (HSs: 88.6%; L2ers: 72.8%) may be due to possible transfer of English intonation: while T4 (high-falling tone) is easy to identify for English speakers in word-final position or isolation since English declarative sentences have falling intonation, T4 in word-initial position is difficult to identify (e.g., Hao 2018). However, without testing all possible T4 combinations, it is difficult to discuss possible transfer on T4. Another explanation provided by Yang and Yang (2019), who found that the T4T3 sequence is ranked the second-most difficult out of the 15 disyllabic possibilities (excluding the T3T3 sequence), is that both T4 and T3 have falling contours.

6.3 HSs vs. L2ers

While HSs and L2ers do not differ significantly, HSs were more native-like than L2ers were in the T3T3, T2T3, and T3T2 conditions. Thus, HSs still show a slight HS advantage over L2ers in tones and T3 sandhi, even though the HS advantage is not as strong as hypothesized. It is likely due to an undesired task effect in the critical condition of T3T3 that led even some NSs to choose the unnatural T3T3 pronunciation and/or due to later acquisition of tone and T3 sandhi in monolingual children than previously reported (see Wong and Strange 2017; Xu Rattanasone et al. 2018; Tang et al. 2019).

Given that T3 sandhi is highly frequent, both HSs and L2ers should have plenty of experience hearing it. However, HSs still performed more native-like

than L2ers. A widely-accepted explanation is that phonetics and phonology are subject to a biologically-determined sensitive (if not critical) period which ends earlier than in other domains (e.g., Granena and Long 2013). Due to decreasing brain plasticity, adult L2ers have difficulty acquiring tones and T3 sandhi, despite frequent occurrences in the input. The finding that HSs are more native-like than L2ers in tones and T3 sandhi is consistent with phonetics/phonology studies with other heritage languages (e.g., Korean: Oh et al. 2003; Spanish: Kim 2020) or other phenomena in heritage Mandarin (e.g., Chang et al. 2011; Yang 2015; Chang and Yao 2016). This is unsurprising since it was found that internationally adopted children who stopped hearing Chinese by age two were still able to retain some unconscious memory of tones (Pierce et al. 2014).

6.4 Pedagogical Implication

In this study, both HSs and L2ers have difficulty correctly identifying sandhied T3, where they were to choose T2 instead of T3 before another T3, and sometimes misidentify half-T3 as T2. Yet, only some L2ers (and not HSs) appear to misidentify half-T3 as T4. It is therefore suggested that instructors can draw learners' attention to sandhied T3 and, especially for L2ers, half-T3 as well. Practice in differentiating T2 and full-T3 is also important, since both have rising tones. As discussed in section 2.5, while "full-T3 first" teaching method might lead learners to confuse T2 and full-T3 (since both have rising tones), the "half-T3 first" does not (since T2 and half-T3 is not easily confusable). As Zhang (2020:173) recommends, "tones should be practiced within tone strings rather than in isolation. This is especially important for the most widely distributed allophone of T3, Half-T3." While students receiving the "half-T3 first" teaching method in He et al. (2016) did not outperform students receiving the "full-T3 first" teaching method on the disyllabic condition (which included T3T1/T3T2/T3T4 and T1T3/T2T3/T3T4, but excluded T3T3), the authors still recommended using the "half-T3 first" teaching method because students receiving this kind of instruction indeed performed better when T3 appears in sentences, and they were better able to generalize their production accuracy from learned words to new words. While more empirical studies might be needed to determine which

teaching method to recommend, drawing learners' attention to both full-T3 and half-T3 is certainly important.

7. Conclusion

The present study found that HSs were indeed more native-like than L2ers in the Tone Identification Task, though their performance was not as native-like as expected. It may be due to task effects that caused even NSs to not score at ceiling on the critical condition (but scored at ceiling on other conditions) and/or later acquisition of tone and T3 sandhi in monolingual children than previously thought (see Wong and Strange 2017; Xu Rattanasone et al. 2018; Tang et al. 2019). One limitation of the present study is that the proficiency is not well controlled using an aural or oral task, like that in Kim (2020). Another limitation is that the task adopted here requires participants to remember the two syllables before identifying the disyllabic tones, but short-term memory is not measured, let alone controlled. Individual differences may have contributed to the not-at-ceiling performance.

Given that tones present consistent difficulty to many L2ers and even some HSs, the present study has potential implications for language pedagogy, which is particularly important given both the growing interest in learning Mandarin and the growing population of Mandarin HSs in college classrooms world-wide. The present study helps identify problem areas facing learners of Mandarin, and provides information about where HSs do or do not face difficulties relative to traditional L2 classroom learners. These findings in turn may potentially inform pedagogical practices: by teaching T3 as a low tone, learners would likely have an easier time differentencing T2 from T3 (which usually surfaces as half-T3 in conversation) and produce or recognize half-T3 under obligatory contexts (see He et al. 2016; Zhang 2017, 2020).

References

- Au, Terry Kit-fong, Leah M. Knightly, Sun-Ah Jun, and Janet S. Oh. 2002. Overhearing a language during childhood. *Psychological Science* 13.3: 238-243.

- Au, Terry Kit-fong, Janet S. Oh, Leah M. Knightly, Sun-Ah Jun, and Laura F. Romo. 2008. Salvaging a childhood language. *Journal of Memory and Language* 58.4: 998-1011.
- Chang, Charles B. 2021. Phonetics and phonology. *The Cambridge Handbook of Heritage Languages and Linguistics*, eds. by Silvina Montrul, and Maria Polinsky, 581-612. Cambridge: Cambridge University Press.
- Chang, Charles B., and Yao Yao. 2016. Toward an understanding of heritage prosody: Acoustic and perceptual properties of tone produced by heritage, native, and second language speakers of Mandarin. *Heritage Language Journal* 13.2: 134-160.
- Chang, Charles B., and Yao Yao. 2019. Production of neutral tone in Mandarin by heritage, native, and second language speakers. *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia*, 2291-2295.
- Chang, Charles B., Yao Yao, Erin F. Haynes, and Russell Rhodes. 2011. Production of phonetic and phonological contrast by heritage speakers of Mandarin. *The Journal of the Acoustical Society of America* 129.6: 3964-3980.
- Chang, Seung-Eun, and Samuel Weiss-Cowie. 2021. Hyper-articulation effects in Korean glides by heritage language learners. *International Journal of Bilingualism* 25.1: 3-20.
- Chang, Yung-hsiang Shawn. 2011. Distinction between mandarin tones 2 and 3 for L1 and L2 Listeners. *Proceedings of the 23rd North American Conference on Chinese Linguistics (NACCL-23)* 1: 84-96.
- Chao, Yuen Ren. 1930. A system of tone letters. *Le Maitre Phonétique* 45: 24-27.
- Chen, Qinghai. 1997. Toward a sequential approach for tonal error analysis. *Journal of Chinese Language Teachers Association* 32: 21-39.
- Chen, Si, Yunjuan He, Ratrete Wayland, Yike Yang, Bei Li, and Chun Wah Yuen. 2019. Mechanisms of tone sandhi rule application by tonal and non-tonal non-native speakers. *Speech Communication* 115: 67-77.
- Duanmu, San. 2000. *The Phonology of Standard Chinese*. Oxford: Oxford University Press.

- Granena, Gisela, and Michael H. Long. 2013. Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research* 29.3: 311-343.
- Hao, Yen-Chen. 2012. Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of phonetics* 40.2: 269-279.
- Hao, Yen-Chen. 2018. Contextual effect in second language perception and production of Mandarin tones. *Speech Communication* 97: 32-42.
- He, Yunjuan, and Ratree Wayland. 2013. Identification of Mandarin coarticulated tones by inexperienced and experienced English learners of Mandarin. *Chinese as a Second Language Research* 2.1: 1-21.
- He, Yunjuan, Qian Wang, and Ratree Wayland. 2016. Effects of different teaching methods on the production of Mandarin tone 3 by English speaking learners. *Chinese as a Second Language. The journal of the Chinese Language Teachers Association, USA* 51.3: 252-265.
- Huang, Yu-hsin. 2006. The role of input in acquisition of tone sandhi rules in Mandarin Chinese. *Proceedings of the 20th Pacific Asia Conference on Language, Information and Computation*, 287-294.
- Jaeger, T. Florian. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59.4: 434-446.
- Jin, Wenhua. 2019. Acquisition of tone sandhis by English speaking learners of Chinese. *Journal of National Council of Less Commonly Taught Languages* 25.1: 67-107.
- Kim, Ji Young. 2019. Heritage speakers' use of prosodic strategies in focus marking in Spanish. *International Journal of Bilingualism* 23.5: 986-1004.
- Kim, Ji Young. 2020. Discrepancy between heritage speakers' use of suprasegmental cues in the perception and production of Spanish lexical stress. *Bilingualism: Language and Cognition* 23.2: 233-250.
- Knightly, Leah M., Sun-Ah Jun, Janet S. Oh, and Terry Kit-fong Au. 2003. Production benefits of childhood overhearing. *The Journal of the Acoustical Society of America* 114.1: 465-474.

- Lee, Yuh-Shiow, Douglas A. Vakoch, and Lee H. Wurm. 1996. Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research* 25.5: 527-542.
- Lenth, Russell, Henrik Singmann, Jonathon Love, Paul Buerkner, and Maxime Herve. 2019. emmeans: Estimated Marginal Means, aka Least-Squares Means (Version 1.3.4).
- Mayberry, Rachel I., and Robert Kluender. 2018. Rethinking the critical period for language: New insights into an old question from American Sign Language. *Bilingualism: Language and Cognition* 21.5: 886-905.
- Li, Xin, Yin Huang, Yunheng Xu, Linxin Yi, Yuming Yuan, and Min Xiang. 2021. Production of tone 3 sandhi by advanced Korean learners of Mandarin. *2021 12th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, 1-5. IEEE.
- Lin, Yu-Jung, and Yu-Yin Hsu. 2018. Whether and how do Mandarin sandhied tone 3 and underlying tone 2 differ? *Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation*, eds. by Stephen Politzer-Ahles, Yu-Yin Hsu, Chu-Ren Huang, and Yao Yao, 376-384. Hong Kong: Association for Computational Linguistics.
- Oh, Janet S., Sun-Ah Jun, Leah M. Knightly, and Terry Kit-fong Au. 2003. Holding on to childhood language memory. *Cognition* 86.3: B53-B64.
- Pelzl, Eric, Ellen F. Lau, Taomei Guo, and Robert DeKeyser. 2019. Advanced second language learners' perception of lexical tone contrasts. *Studies in Second Language Acquisition* 41.1: 59-86.
- Peng, Shu-Hui. 2000. Lexical versus 'phonological' representations of Mandarin sandhi tones. *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, eds. by Michael B. Broe, and Janet B. Pierrehumbert, 152-167. Cambridge: Cambridge University Press.
- Pierce, Lara J., Denise Klein, Jen-Kai Chen, Audrey Delcenserie, and Fred Genesee. 2014. Mapping the unconscious maintenance of a lost first language. *Proceedings of the National Academy of Sciences* 111.48: 17314-17319.
- R Core Team. 2019. R: A language and environment for statistical computing.

- Austria, Vienna: R Foundation for Statistical Computing. Accessed online, July 26, 2020. <https://www.R-project.org/>
- Shih, Chinlin. 1986. *The Prosodic Domain of Tone Sandhi in Chinese*. San Diego: University of California, San Diego Ph. D. dissertation.
- Speer, Shari R., Chinlin Shih, and Maria L. Slowiaczek. 1989. Prosodic structure in language understanding: evidence from tone sandhi in Mandarin. *Language and Speech* 32.4: 337-354.
- Tai, James H.-Y. 1978. *Phonological Changes in Modern Standard Chinese in the People's Republic of China Since 1949*. Washington, DC: Office of Research, United States Information Agency.
- Tang, Ping, Ivan Yuen, Nan Xu Rattanasone, Liqun Gao, and Katherine Demuth. 2019. The acquisition of phonological alternations: The case of the Mandarin tone sandhi process. *Applied Psycholinguistics* 40.6: 1495-1526.
- Tsao, Feng-Ming. 2016. Speech perception in normally developing infants and children. *Encyclopedia of Chinese Language and Linguistics*, eds. by Rint Sybesma, Wolfgang Behr, Yueguo Gu, Zev Handel, C.-T. James Huang, and James Myers. Leiden: Brill. Accessed online, February 17, 2017. http://dx.doi.org/10.1163/2210-7363_ecll_COM_000208
- Tse, Holman. 2016. Variation and change in Toronto heritage Cantonese: An analysis of two monophthongs across two generations. *Asia-Pacific Language Variation* 2.2: 124-156.
- Tsay, Jane. 2016. Acquisition of tone, L1. *Encyclopedia of Chinese Language and Linguistics*, eds. by Rint Sybesma, Wolfgang Behr, Yueguo Gu, Zev Handel, C.-T. James Huang, and James Myers. Leiden: Brill. Accessed online, February 17, 2017. http://dx.doi.org/10.1163/2210-7363_ecll_COM_000106
- Tsukada, Kimiko, Hui Ling Xu, and Nan Xu Rattanasone. 2015. The perception of Mandarin lexical tones by listeners from different linguistic backgrounds. *Chinese as a Second Language Research* 4.2: 141-161.
- Tu, Jung-Yueh, Janice Wong W.-S., and Jih-Ho Cha. 2017. Trisyllabic tone 3 sandhi patterns in Mandarin produced by Cantonese speakers. *Proceedings of Interspeech 2017*, 3177-3180.
- Tu, Jung-Yueh, and Yu-Fu Chien. 2020. The processing of Mandarin Chinese

- tonal alternations in contexts: An eye-tracking study. *ICASSP 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 6244-6248. IEEE.
- Wang, Yue, Joan A. Sereno, and Allard Jongman. 2006. L2 acquisition and processing of Mandarin tones. *The Handbook of East Asian Psycholinguistics*, eds. by Ping Li, Li Hai Tan, Elizabeth Bates, and Ovid J. L. Tzeng, 250-256. Cambridge: Cambridge University Press.
- Wang, Yue, Allard Jongman, and Joan A. Sereno. 2016. Acquisition of tone, L2. entry. *Encyclopedia of Chinese Language and Linguistics*, eds. by Rint Sybesma, Wolfgang Behr, Yueguo Gu, Zev Handel, C.-T. James Huang, and James Myers. Leiden: Brill. Accessed online, February 17, 2017. http://dx.doi.org/10.1163/2210-7363_ecll_COM_000105
- Wong, Pui-san, and Winifred Strange. 2017. Phonetic complexity affects children's Mandarin tone production accuracy in disyllabic words: A perceptual study. *PloS One* 12.8: e0182337.
- Xu Rattanasone, Nan, Ping Tang, Ivan Yuen, Liqun Gao, and Katherine Demuth. 2018. Five-year-olds' acoustic realization of Mandarin tone sandhi and lexical tones in context are not yet fully adult-like. *Frontiers in Psychology* 9: 817.
- Yang, Bei. 2015. *Perception and Production of Mandarin Tones by Native Speakers and L2 Learners*. Berlin: Springer Berlin Heidelberg.
- Yang, Bei, and Nuoyi Yang. 2019. Development of disyllabic tones in different learning contexts. *International Review of Applied Linguistics in Language Teaching* 57.2: 205-233.
- Yang, Chunsheng. 2016. *The Acquisition of L2 Mandarin Prosody: From Experimental Studies to Pedagogical Practice (Vol. 1)*. Amsterdam: John Benjamins Publishing Company.
- Yang, Chunsheng, and Wenhua Jin. 2018. Chinese as a second language pronunciation teaching survey. *Journal of the National Council of Less Commonly Taught Languages* 23: 153-189.
- Yuan, Jiahong, and Yiya Chen. 2014. 3rd tone sandhi in standard Chinese: A corpus approach. *Journal of Chinese Linguistics* 42.1: 218-237.

- Zhang, Caicai, and Gang Peng. 2013. Productivity of Mandarin third tone sandhi: A wug test. *Eastward Flows the Great River: Festschrift in Honor of Prof. William SY. Wang on His 80th Birthday*, eds. by Gang Peng, and Feng Shi, 255-282. Hong Kong: City University of Hong Kong Press.
- Zhang, Hang. 2017. The effect of theoretical assumptions on pedagogical methods: A case study of second language Chinese tones. *International Journal of Applied Linguistics* 27.2: 363-382.
- Zhang, Hang. 2018a. Current trends in research of Chinese sound acquisition. *The Routledge Handbook of Chinese Second Language Acquisition*, ed. by Chuanren Ke, 217-233. New York: Routledge.
- Zhang, Hang. 2018b. *Second Language Acquisition of Mandarin Chinese Tones: Beyond First-language Transfer*. Leiden: Brill.
- Zhang, Hang. 2020. Teaching Chinese Tones. *The Routledge Handbook of Chinese Language Teaching*, eds. by Chris Shei, Monica E. McLellan Zikpi, and Der-lin Chao, 166-179. New York: Routledge.
- Zhang, Jie, and Yuwen Lai. 2010. Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology*, 153-201.
- Zhu, Hua. 2002. *Phonological Development in Specific Contexts: Studies of Chinese-speaking Children (Vol. 3)*. Clevedon: Multilingual Matters.
- Zhu, Hua. 2016. Acquisition of phonology, L1. *Encyclopedia of Chinese Language and Linguistics*, eds. by Rint Sybesma, Wolfgang Behr, Yueguo Gu, Zev Handel, C.-T. James Huang, and James Myers, 00-00. Leiden: Brill. Accessed online, February 17, 2017. http://dx.doi.org/10.1163/2210-7363_ecll_COM_000105

[审查：2021.10.1 修改：2021.11.4 接受：2021.12.23]

華語文教學研究

陳仲妤

Chung-yu CHEN

11651 臺北市文山區指南路二段 64 號

國立政治大學華語文教學碩博士學位學程

Master's & Doctor's Program in Teaching Chinese as a Second Language

National Chengchi University

No. 64, Sec. 2, ZhiNan Rd., Wenshan District, Taipei City 11651, Taiwan

chungyu@nccu.edu.tw

石基琳

Chilin SHIH

Department of East Asian Languages and Cultures, Department of Linguistics

University of Illinois at Urbana-Champaign

707 S. Mathews Avenue, Urbana, Illinois 61801, U.S.A.

cls@illinois.edu

Appendix: Blocks and Items in the Tone Identification Task

	Block 1	Block 2	Block 3	Block 4	Block 5
T1-T3	<i>yi1ma3</i>	<i>da1yi3</i>	<i>ke1da3</i>	<i>yi1ke3</i>	<i>da1you3</i>
T2-T3	<i>da2you3</i>	<i>yi2ma3</i>	<i>da2yi3</i>	<i>ke2da3</i>	<i>yi2ke3</i>
T3-T3	<i>yi3ke3</i>	<i>da3you3</i>	<i>yi3ma3</i>	<i>da3yi3</i>	<i>ke3da3</i>
T4-T3	<i>ke4da3</i>	<i>yi4ke3</i>	<i>da4you3</i>	<i>yi4ma3</i>	<i>da4yi3</i>
T3-T1	<i>da3yi1</i>	<i>ke3da1</i>	<i>yi3ke1</i>	<i>da3you1</i>	<i>yi3ma1</i>
T3-T2	<i>yi3ma2</i>	<i>da3yi2</i>	<i>ke3da2</i>	<i>yi3ke2</i>	<i>da3you2</i>
T3-T4	<i>da3you4</i>	<i>yi3ma4</i>	<i>da3yi4</i>	<i>ke3da4</i>	<i>yi3ke4</i>

華語為繼承語和第二語言習得： 以三聲變調感知為例

陳仲好

國立政治大學華語文教學碩博士學位學程

石基琳

伊利諾大學香檳校區東亞語言與文化系、語言學系

摘要

本文旨在檢視以英語為強勢語言的華裔繼承語使用者 (heritage speakers) 和成人後方學習華語為第二語言之學習者 (下稱二語者) 之三聲變調感知。當兩個三聲連讀，第一個三聲變為二聲；當三聲在非三聲之前，該三聲變成半三聲。42 位華語母語人士、21 位繼承語使用者和 25 位二語者參與了聲調辨識實驗，受試者在聽了兩個單音節之後，須從四個音檔中選出母語人士如何將剛聽過的兩個單音節連讀成一個雙音節。雖然有些統計結果未達顯著，整體趨勢顯示：(1) 與二語者相比，繼承語使用者有些許優勢 (與母語者更加相似)；(2) 跟其他音調組合相較，繼承語使用者在辨識三聲變調和半三聲有些許困難，但二語者似乎對所有雙音節組合皆有困難。

關鍵詞：三聲變調 以英語為母語的二語者 以英語為強勢語言的繼承語習得者 半三聲 感知