華語文教學研究 20.3: 57-88, 2023

# An Empirical Study on Teaching Half-T3 First in L2 Mandarin*

**Yun-Chih HUANG** **Yu-Hsin HUANG**
**International Graduate Program of Teaching Chinese as a Second Language**
**Tunghai University**

## Abstract

This study investigates the teaching and learning of Mandarin Chinese tones, with a specific focus on pedagogical aspects related to the third tone (T3), which are developed from different theoretical assumptions concerning the default form of T3. The traditional teaching method asserts that the full-T3 (with a dipping contour, [214]) is the standard form, while the innovative approach posits the half-T3 (with a low falling contour, [21]) as the standard. Since no experimental study has examined the impact of these two teaching methods, a training study was conducted to compare their effectiveness and determine which one might be more beneficial for L2 learners. Thirty-two non-tonal speakers participated in this study and a two-stage production test (comprising an immediate posttest and a delayed posttest) was employed to assess the participants' T3 performance. The findings indicate that the Half-T3 First (HT3F) teaching method generally enhances the tone production skill of L2 learners across most tone combinations. However, in the case of the T3-T3 combination, a significant decrease in pre-T3 accuracy and subpar performance in

* We thank three anonymous reviewers for providing valuable and helpful comments on an earlier version of the article.

phrase-final T3 were observed. This unexpected outcome was attributed to the general principle of faithfulness within the framework of Optimality Theory, suggesting that the HT3F may encounter constraint violations.

**Keywords:** Computational Complexity Hypothesis, Full-T3 First, Half-T3 First, Optimality Theory, Tonal Markedness Scale

## 1. Introduction

Mandarin Chinese is a tonal language, and the accurate pronunciation of tones is crucial for conveying the intended meaning in words. Mispronouncing tones can lead to misunderstandings during communication. Several studies have explored the challenges faced by non-native learners, such as American students, in accurately producing Mandarin tones (e.g., Miracle 1989; Shen 1989; Chen 1997; Wang et al. 2003; He 2014; Yang 2016). These studies have highlighted the difficulties that learners may encounter in mastering the tonal distinctions of the language.

The third tone (T3) in Mandarin is traditionally analyzed with three distinct allophones, as represented in Table 1: (a) full-T3, characterized by a low-dipping contour, notated as [214][1], (2) pre-T3, featuring a mid-rising contour, denoted as [35], and (3) half-T3, displaying a low-falling contour, represented as [21]. The full-T3 variant is typically pronounced when the syllable stands alone or when it appears at the end of a phrase. The pre-T3 variant is used when T3 precedes another T3 syllable in disyllabic combinations, while the half-T3 variant is employed before the first tone (T1), the second tone (T2), or the fourth tone (T4) in disyllabic combinations.

Table 1: The “Full-T3 First (FT3F)” Method (Zhang 2018:109)

| T3 variant | Pitch value | Tone contour | Environment of occurrence |
|---|---|---|---|
| Full-T3 | [214] | Low-dipping | In isolation or at phrase-final |

[1] Chao (1930) invented a numeric tonal notation system in which 1 represents the lowest tone and 5 signifies the highest tone.

| | | | |
|---|---|---|---|
| | | | positions |
| Pre-T3 | [35] | Mid-rising | Preceding T3 in disyllabic combinations |
| Half-T3 | [21] | Low-falling | Preceding T1, T2, or T4 in disyllabic combinations |

As outlined in Zhang's (2018) scholarly research, the prevailing pedagogical approach for teaching T3 allophones in the United States, the People's Republic of China, and several other countries is the Full-T3 First method (abbreviated as FT3F) (Zhang 2018:86). This pedagogical approach involves initially instructing the full-T3 [214] as the base form of T3, followed by the introduction of pre-T3 [35] and finally, half-T3 [21], in the sequence outlined in Table 1. The terms "Full-T3 First", although frequently used in the literature and adopted here, can be somewhat misleading. While the prevailing assumption is that full-T3 serves as the underlying form, with half-T3 derived from it, this study aligns with Zhang (2018) in defining the Full-T3 First method as an approach that introduces L2 learners to full-T3 as the primary reference point (hence, Full-T3 First).

Zhang underscores that, among the three T3 allophones, the most commonly employed half-T3 variant[2] is often disregarded or receives only cursory attention during classroom instruction (Zhang 2018:86). A similar observation has been made by Tsung (1987) and Chen and Shih (2021), who note that instructed learners are generally familiar with the pre-T3 sandhi rule but have limited awareness of the half-T3 rule, despite its frequent usage. Upon closer examination, this study has discovered that the two widely adopted teaching materials in Taiwan completely omit any mention of half-T3 [21]. Please refer to appendices A and B for the instructions regarding the change of the third tone in *Practical Audio-Visual Chinese* (National Taiwan Normal University Mandarin Training Center 1999) and in *A Course in Contemporary Chinese* (National Taiwan Normal University Mandarin Training Center 2021). Therefore, it is imperative to reevaluate the current prevailing pedagogical method, which

[2] Duanmu (2000) suggests that the majority of native Chinese speakers tend to pronounce half-T3 [21] instead of full-T3 [214] when producing T3 in isolation or at the phrase-final position, establishing half-T3 as the most commonly employed allophone of T3.

prioritizes full-T3 and pre-T3 while neglecting half-T3.

This study delves into the effectiveness of teaching the Half-T3 First method, hereafter referred to as HT3F (as depicted in Table 2), to individuals who are native speakers of English or other non-tonal languages and have had no prior exposure to Mandarin. The aim is to gather data that can serve as a foundation for an alternative instructional approach. This alternative method is anticipated to reduce the difficulty level for learners when they initially encounter T3 sandhi rules during the early stages of Mandarin language acquisition. If this modified approach proves to be more effective, it will provide educators with valuable insights for designing courses that facilitate improved mastery of T3, widely recognized as the most challenging tone in Mandarin.

Table 2: The “Half-T3 First (HT3F)” Method Adopted from Zhang (2018:109)

| T3 variant | Pitch value | Tone contour | Environment of occurrence |
|---|---|---|---|
| Half-T3 | [21] | Low-falling | In isolation or at phrase-final positions; preceding T1, T2, or T4 in disyllabic combinations |
| Pre-T3 | [35] | Mid-rising | Preceding T3 in disyllabic combinations |

## 2. Literature Review

This section reviews pedagogical and theoretical research on the acquisition of Mandarin T3. Section 2.1 and 2.2 review research on the FT3F and the HT3F methods, respectively. Section 2.3 presents the theoretical frameworks underpinning this study: the Tonal Markedness Scale (Cheng 1973; Ohala 1978; Hyman and Van Bik 2004) within the framework of Optimality Theory (McCarthy and Prince 1993; 1995; Prince and Smolensky 2004), and the Computational Complexity Hypothesis (Jakubowicz 2011; Prévost et al. 2014; Yuan, 2015). Section 2.4 provides an interim summary.

### 2.1 The FT3F Method

The FT3F approach to teaching Mandarin Chinese T3 considers full-T3 [214] as the fundamental or base form of T3. In the FT3F approach, instruction for L2 learners begins with full-T3 as the primary reference point, hence the term "Full-T3 First". Learners are subsequently taught to memorize the pre-T3 sandhi rule to generate the [35] allophone and then the half-T3 sandhi rule to generate the [21] allophone. Theorizing [214] as T3's base form has had a significant influence on second-language pedagogy in various ways. This assumption aligns with the tone mark for T3, which is represented by the falling-rising shape of the T3 phonetic transcription used in the *Zhuyin* system and teaching materials.

However, as early as 1973, Chen observed a "tendency of students who treated full-T3 as the norm form of T3, using full-T3 in all environments" (Chen 1973:146). Chen also pointed out that learners from non-tonal language backgrounds fail to recognize half-T3 as an allophone of T3. Zhang (2018) conducted experiments and discovered a notably high error rate in the use of half-T3 among the three T3 allophones, particularly involving the overproduction of full-T3 in situations where half-T3 should be used (Zhang 2018:104). In theory, substituting an allophone for another allophone should not lead to a change in meaning (e.g., substituting full-T3 for half-T3), as the substitution does not alter the phoneme's underlying identity. However, when it comes to Chinese tones, mispronouncing the allophones of T3 can result in not only poor tonal performance but also potential misunderstandings. This happens because the mispronunciation of allophones can trigger a chain reaction that affects adjacent tones due to tonal coarticulation effects in connected speech (Xu 1997). This coarticulation effect may decrease the intelligibility of the phonological identities of L2 tones. Therefore, it is crucial for learners to produce the correct T3 allophone in specific linguistic environments to achieve native-like spoken Chinese and minimize potential communication difficulties.

As observed in the literature, T3 has consistently been considered one of the most challenging tones for L2 learners, affecting both perception and production. Numerous studies have documented a high error rate among L2 learners when it comes to T3, and they often acquire proficiency in this tone relatively late (e.g., Chen 1997; Winke 2007; Shi 2007; Zhang 2014). Despite the abundance of L2 studies investigating T3 acquisition, there is a notable lack of connections between phonological theories and pedagogical approaches. The majority of previous research treats T3 variations as a single entity, examining them collectively without distinguishing between the variants or the specific contexts in which T3 appears, leaving unclear which variation(s) or environment(s) pose the greatest difficulty. This study aims to bridge these gaps by conducting a novel experiment and analysis that assess each of T3's variants and the contexts in which it occurs individually. This approach will provide insight into the distinct challenges encountered by non-native learners as they strive to master Mandarin Chinese T3.

## 2.2 The HT3F Method

Given the high error rate associated with half-T3 and the tendency to overproduce full-T3, as discussed in Section 2.1, Tsung (1987:96) proposed a shift in the focus of teaching Mandarin T3 towards half-T3. This suggestion arises despite the widespread use of the FT3F approach as the most commonly employed teaching method. Additionally, Zhang (2018:109) advocated for the HT3F method, with a particular emphasis on the significance of teaching half-T3 as the base form. Zhang's HT3F method introduces two variants of T3 to L2 learners: (1) half-T3 [21] for cases where T3 is produced in isolation, at phrase-final positions, or preceding T1, T2, or T4, and (2) pre-T3 [35] when T3 is produced before another T3 in disyllabic combinations. Notably, Zhang's HT3F method excludes the full-T3 variant because he follows Hyman (2007) and Chen and Xu (2006), regarding [214] as simply a natural intonation form. According to Zhang, "there is no need to emphasize the Full-T3 form because the 'bounce' effect for the final rise pitch is a natural result of releasing the laryngeal muscles, which have contracted for low pitches (i.e., the base tone Half-T3), to a relaxed state" (Zhang 2018:108).

It is believed that adopting the HT3F method can reduce the cognitive load on learners, as it necessitates the acquisition of only two T3 variants and one sandhi rule (as presented in Table 2 in Section 1).

Despite its promising efficacy, the HT3F method has encountered significant resistance due to its contradiction with the widely accepted notion of full-T3 [214] as the underlying form of T3. It is common scholarly convention to define a tone in isolation as its fundamental value (Mei 1977:253), which seemingly establishes full-T3 [214] the primary value of T3 and positions half-T3 [21] as secondary. However, Mei (1977) delved into the historical and contemporary usage of T3 and asserted, “By adopting [21] as the basic value of the third tone, the value of the third tone did not change between the 16th century and the present” (Mei 1977:253). According to Mei (1977), [21] has been in use since the 16th century, while [214] represents a diachronic variant that has emerged only recently in contemporary Chinese (Mei 1977:255). Zhang (2018) similarly argued, “There is cause to think that half-T3, not full-T3, is the underlying form, given that half-T3 comes before T1, T2, T4, whereas full-T3 only appears on its own” (Zhang 2018:84).

Research conducted on native speakers also suggests that half-T3 might be a more accurate representation of the underlying form of T3. Even in phrase-final positions, native speakers tend to produce half-T3 rather than full-T3. Duanmu (2000:238) investigated the contemporary usage of T3 among six native Chinese speakers, asking them to produce 16 phrases, each containing T3 in the final position. Among these speakers, five consistently pronounced T3 as [21]. Notably, only one individual, who had undergone an entrance exam for the Beijing Broadcasting Institute with aspirations of becoming a broadcaster, pronounced four out of the 16 phrase-final T3s as [214]. This difference can be attributed to the heightened attention to pronunciation exhibited by the aspiring broadcaster.

## 2.3 Theoretical Framework: The Tonal Markedness Scale within the framework of Optimality Theory and the Computational Complexity Hypothesis

Past research on the L2 acquisition order of individual tones generally aligns with findings from studies on L1 acquisition (Li and Thompson 1977; Zhu and Dodd 2000). This parallel suggests the presence of Universal Grammar (UG) (Chomsky 1965; 1981) in L2 grammars. Over the past few decades, research on L2 phonological acquisition has provided evidence indicating that interlanguage phonologies are influenced not only by a speaker’s L1, but also constrained by UG (Broselow et al. 1998; Major 2001).

In this Section, the involvement of one universal phonological principle and one computational principle in the L2 acquisition of Mandarin tones will be considered: (1) the Tonal Markedness Scale (TMS) (Cheng 1973; Ohala 1978; Hyman and VanBik 2004) that is phonetically grounded and formalized in the theoretical framework of Optimality Theory (OT) (McCarthy and Prince 1993; 1995; Prince and Smolensky 2004), and (2) the Computational Complexity Hypothesis (CCH) (Jakubowicz 2011; Prévost et al. 2014; Yuan 2015).

In OT, UG is conceptualized as a universal set of constraints inherent to all languages, and the process of L1 acquisition is involves re-ranking this universal set of constraints. These constraints are assumed universal, with differences between languages arising from the way these universal constraints are ranked differently. Consequently, OT grammars are comprised of a set of ranked constraints that determine the optimal output corresponding to an input string. To evaluate what this optimal output might be, two types of constraints are employed: (1) markedness constraints and (2) faithfulness constraints[3].

In the context of Mandarin tones, the markedness constraint is often represented by the well-established Tonal Markedness Scale (TMS) (Cheng 1973; Ohala 1978; Hyman and VanBik 2004), which defines "markedness" as the level of physical effort required to articulate sounds (Yip 2002). According to the TMS, the hierarchy of tone markedness is as follows: Rising tones are more marked than falling tones, and falling tones are more marked than level tones. Consequently, when considering T3 variants based on the TMS, their markedness ranking from high to low is as follows: full-T3 [214] > pre-T3 [35] > half-T3 [21]. In other words, within the framework of the TMS, the base form of T3 in the FT3F method (i.e., [214]) is inherently more marked than that in the HT3F method (i.e., [21]).

Another framework employed in this study is the Computational Complexity Hypothesis (CCH) (Jakubowicz 2011; Prévost et al. 2014; Yuan 2015), which posits that language structures demanding less computational effort are typically acquired earlier in development than structures requiring greater computational

[3] The faithfulness constraints will be reviewed in detail in Section 5.2.

load. A key concept within the CCH is the derivational complexity matrix (DCM) introduced by Jakubowicz (2011). According to the DCM, a process involving merging *n* times gives rise to a less complex derivation than one involving merging (*n* + 1) times, and a process involving merging α gives rise to a less complex derivation than one involving merging α + ß. Jakubowicz (2011) suggests that the DCM "applies to different conditions of language acquisition (L1, L2, SLI, etc.) and to adult processing as well" (Jakubowicz 2011:340).

Extending this DCM to the L2 acquisition of tones, we can assume that when computing tonal complexity, such as T3 variants, learners are sensitive to the number of steps required to generate a specific T3 variant in various disyllabic contexts of usage (please refer to Table 1 and 2). Learners are assumed to start with less complex computations (e.g., the base form, which is the first variant in T3 acquisition) and, all other things being equal, acquire pre-T3 that is derived from the base form. In other words, learners may initially prefer variants that require only one step (e.g., the base form) rather than two steps (e.g., pre-T3 derived from the base form), or three steps (e.g., full-T3 derived by excluding contexts where pre-T3 can occur).

This approach assigns distinct levels of complexity to the two T3 pedagogical methods. For example, the FT3F method entails a three-step process for learners to acquire half-T3. They initially learn full-T3 as the base form, then proceed to acquire pre-T3, and finally master half-T3, as outlined in Table 1. In contrast, the HT3F method simplifies the learning process considerably by allowing learners to acquire half-T3 in just one step—the base form, as illustrated in Table 2. Table 3 summarizes the similarities and differences between the two pedagogies in terms of "complexity" and "markedness".

Table 3: Similarities and Differences between FT3F and HT3F within the CCH and the TMS

| Method | Step 1 (base form) | Step 2 | Step 3 |
| --- | --- | --- | --- |
| FT3F | Full-T3 at phrase-final positions ([214], most marked) | Pre-T3 when preceding T3 ([35], second least marked) | Half-T3 when preceding T1, T2, or T4 ([21], least marked) |
| HT3F | Half-T3 at phrase-final positions or when preceding T1, T2, or T4 ([21], least marked) | Pre-T3 when preceding T3 ([35], second least marked) | NA |

### 2.4 Interim Summary

Up to this point, we have discussed the historical usage of T3, which has predominantly been pronounced as [21] from the 16th century to the present day. Building upon this historical context and considering the frequent usage of half-T3 [21] observed among native speakers by Duanmu (2000), Zhang (2018) proposed a shift from the FT3F method to the HT3F. In the HT3F approach, only two T3 variants are taught, with half-T3 serving as the base form of T3.

Following the principles of the TMS and the CCH, this study regards the HT3F method as a less marked and less complex approach for acquiring T3 variants. It suggests that commencing instruction with the HT3F method could potentially lead to more effective T3 acquisition for learners, for two key reasons.

Firstly, within the CCH framework, the FT3F method requires learners to go through three steps (variants) to fully acquire T3 pronunciation in all disyllabic positions, whereas the HT3F method streamlines this process to just two steps.

Secondly, within the TMS framework, the FT3F method initiates with the most marked variant [214], while the HT3F method starts with the least marked variant [21].

Consequently, learners following the HT3F method, which should facilitate a faster and more efficient acquisition of T3 variants, are expected to encounter less difficulty compared to those in the FT3F method. To assess the impact of these two teaching methods, we conducted a training study, the details of which will be elaborated upon in Section 3.

## 3. Methodology

To compare the effectiveness of the FT3F method and the HT3F method in teaching T3 to non-tonal language speakers, this study involves a total of 32 participants, who are divided into a control group (CG) and an experimental group (EG). The research questions and predictions based on the TMS and the CCH are presented in Section 3.1. Information about the participants and the procedures for the study are outlined in Section 3.2 and 3.3 respectively.

### 3.1 Research Questions and Predictions

To investigate the effects of the FT3F and HT3F methods on learners' accuracy in producing different T3 variants and to explore their implications for the CCH and the TMS, a training study and two post-treatment production tests, referred to as posttests, were conducted to compare the two pedagogical approaches. These two posttests consisted of an immediate test and a delayed test, which were used to assess the retention of the two pedagogical methods.

This study employed the HT3F and the FT3F method on two separate groups of learners to address the following research questions:

1. Which group demonstrated a higher rate of accuracy in producing pre-T3 during the immediate and delayed posttests?
2. Which group exhibited a higher rate of accuracy in producing phrase-final T3 during the immediate and delayed posttests?
3. Which group displayed a higher rate of accuracy in producing T3 preceding T1, T2, or T4 during the immediate and delayed posttests?
4. Does this study provide support for the CCH and the TMS?

The study aimed to answer these questions to gain insights into the effectiveness of the two teaching methods and their alignment with theoretical frameworks such as the CCH and the TMS.

In this study, 32 learners were divided into two groups: a control group (CG) and an experimental group (EG). The CG followed the traditional FT3F method, whereas the EG used the HT3F method. Given the varying levels of complexity experienced by these two groups, with the CG tasked with acquiring three T3 variants and two sandhi rules while the EG only needed to learn two variants and one sandhi rule, along with the recognized tonal markedness of T3 variants, the study proposed a hypothesis suggesting that EG would achieve a significantly higher rate of accurate T3 pronunciation across all tone combinations. That is, following the principles of the TMS and the CCH, this study regards the HT3F method as a less marked and less complex approach for acquiring every T3 variant (i.e., the half-T3 variant at phrase-final positions and when preceding T1, T2, or T4, as well as the pre-T3 variant when preceding another T3).

### 3.2 Participants

This primary objective of this study was to evaluate the effectiveness of two T3 pedagogical methods. In order to minimize the potential impact of prior familiarity with tone languages on the outcomes, a total of 32 participants were selected. The EG had 8 males and 8 females, with an age range of 18-30 and an average age of 20.5, while the CG had 9 males and 7 females ranging from 18 to 34 years old, with an average age of 20.8. These participants were native speakers of non-tonal languages, including Czech (2 participants), English (2 participants), Indonesian (23 participants), Italian (1 participant), Portuguese (1 participant), Slovak (1 participant), Spanish (1 participant), and Turkish (1 participant), and had not received any formal instruction in any tonal languages.

### 3.3 Procedures

This research was conducted in multiple stages to assess the effectiveness of the FT3F and HT3F methods in teaching T3 to non-tonal language speakers. Prior to commencing the experiment, participants received an email notification one day before the experiment. This email included details about the date and time of their individual online meeting via Google Meet, as well as information about the necessary equipment for the study. During the experiment, participants started by watching a treatment video via Google Meet. After viewing the treatment video, participants watched the posttest video, and their oral productions were recorded using screen recording on the researcher’s computer to capture their responses.

One week following the treatment, participants watched the same posttest video again and their oral productions were recorded. This second round of testing served as the delayed production test, assessing the retention of tones. The final stage of the experiment study involved two Mandarin teachers analyzing the participants' recorded performances to evaluate their production of various T3 allophones.

To ensure that participants remained attentive during the home-based portions of the experiment, measures were implemented. Participants were informed that they could watch the video only once, emphasizing the importance of focused viewing and ensuring their active engagement with the content.

### 3.3.1 Treatment and Materials

The treatment video provided participants with instruction on Mandarin tones via YouTube. Both groups received similar instruction for T1, T2, and T4, but they differed in the instruction of T3. The sequencing of tone instruction was as follows:

CG: T2 >> T4 >> T1 >> T3 in the FT3F method

EG: T2 >> T4 >> T1 >> T3 in the HT3F method

This experiment primarily focused on tonal production rather than articulatory production. Consequently, both the treatment and production tests exclusively utilized the vowel [a] to elicit T1, T2, T4, and the T3 variants. The duration of the treatment video for the CG is 4 minutes and 42 seconds, whereas for the EG, it is 4 minutes and 38 seconds.

Following a video demonstration of monosyllabic instances of T1, T2, and T4, participants were instructed on how to pronounce disyllabic combinations of T1, T2, and T4. Subsequently, the video provided guidance on the T3 variants. Visual aids in the form of images were incorporated to assist in remembering tone contours (refer to Table 4 and appendix C). The selection of images assigned to Mandarin tones underwent two pilot tests, each involving five non-tonal language speakers. Originally, for T2, the image depicted a rocket shooting into the sky, which mirrored the tone contour of T2. Similarly, for T4, the original image portrayed a tilted hot air balloon about to land, aligning with the T4 contour. However, participants in the first pilot study found it challenging to associate these images with the corresponding tones. The current images assigned to T2 and T4 were suggested by one pilot participant and further tested. The current images for T2 and T4 now consist of a question mark and an exclamation mark, which all participants in the second pilot study found easier to associate with the sounds based on their cognitive understanding. Question marks typically signify rising tones, while exclamation marks signify attention, creating stronger associations with falling tones for learners. It's worth noting that the EG using the HT3F method was not presented with an image or any instruction of full-T3.

Additionally, when two T3s occur consecutively, the first syllable changes to the second tone, which is the pre-T3 variant. Consequently, the image of T2 was also used to represent pre-T3.

Table 4: Images Assigned to Mandarin Tones

| Tone | T2/Pre-T3 | T4 | T1 | Full-T3 | Half-T3 |
|---|---|---|---|---|---|
| Image |  |  | |  | |

3.3.2 Two-stage post-treatment production tests

Both groups of participants underwent two posttests. The first posttest was conducted immediately after they watched the treatment video, while the second posttest occurred a week later. During these posttests, participants watched a video on YouTube that prompted them to produce sounds in response to on-screen images, and their oral output was recorded. On the day before the experiment, participants received a message with the following content: "To ensure a successful session, please ensure and remember the following:

1. Your internet connection is stable.
2. Your computer's microphone and camera are functioning properly, and screen sharing is enabled."

Participation in the experiment required a stable internet connection, and a properly functioning computer camera and microphone. The use of headphones was optional. The researcher used screen recording on the computer as the recording tool.

Before starting the posttest, participants were presented with a set of five words, including two monosyllabic words and three disyllabic words. This served to familiarize participants with the pace of the video and acted as a trial run to ensure consistency and prepare them for the subsequent posttest.

A prerequisite test was incorporated into the immediate posttest. This prerequisite test assessed the four isolated tones (T1, T2, T3[4] and T4), with each tone tested four times, totaling 16 tokens. To proceed to the production test, participants needed to accurately produce each tone at least three times. Prior to consenting to participate in the experiment, participants were informed that failing to pass this prerequisite test would result in their exclusion from the study, with partial compensation provided. This prerequisite test served as a strong incentive for participants to engage seriously with the treatment video. Three individuals who did not pass the prerequisite test were excluded from participating in the immediate test before its official commencement. Eventually, 32 participants in the study met this criterion, thus helping to mitigate the potential impact of varying levels of concentration among participants on the experimental results.

Delayed posttests were employed to assess retention, enabling researchers to evaluate participants' ability to retain the knowledge they had acquired over time. These delayed posttests were administered one week after the immediate posttest. Participants were presented with the same video as in the immediate test, excluding the prerequisite test. One-on-one online meetings were organized via Google Meet to conduct both posttests.

Each posttest comprised a total of 42 tokens, with each token testing a disyllabic combination. In these posttests, participants were instructed to produce 28 disyllabic words as targeted stimuli involving T3, which are listed in Appendix D. Additionally, there were 14 disyllabic distractors encompassing four different tone combinations (i.e., T2+T4, T4+T2, T1+T2, T2+T1). The sequence of the disyllabic words in the tests was randomized to prevent predictability. Participants were given a time limit of 5 seconds to respond to each disyllabic

[4] For the CG, T3 in isolation is pronounced as full-T3. In contrast, for the EG, T3 in isolation is pronounced as half-T3.

prompt during the posttests.

### 3.3.3 Assessing Procedures

The participants' pronunciations of T3 in the 28 disyllabic combinations involving T3 in each of the two posttests were evaluated through human auditory perception. This assessment was conducted by two native Mandarin speakers, both of whom had substantial experience as teachers of Chinese as a second language.

Before grading, the two Mandarin teachers listened to the audio files of the initial four monosyllabic tones and view the corresponding images from the treatment video. This was done to ensure that both assessors' judgments were in agreement.

The two Mandarin teachers were instructed to mark on the grading sheet whether a specified tone was pronounced correctly or incorrectly by ticking "correct" or "incorrect". No additional markings were required. The inter-rater reliability between the two assessors was 100%. The incorrect pronunciations were later analyzed by the researchers using Praat to identify possible tone contours.

Following the assessments, statistical analyses were conducted to determine the rate at which each participant accurately produced T3 across various tone combinations.

## 4. Results

The target T3 variants were categorized into three different contexts of T3 occurrence, which were coded as follows: pre-T3 (coded as A), phrase-final T3 (coded as B), and T3 preceding T1, T2, or T4 (coded as C).

The statistical analyses conducted in this study were nonparametric. Since both the CG and the EG in this study each consist of only 16 participants, which is less than 30, they fall into the category of a small sample size. In Krithikadatta's (2014) research, it is emphasized that sample size plays a significant role in data distribution. The distribution chart deviates from the typical bell curve shape when the sample size is 10, 15, or 20. However, as the sample size grows to 25, the distribution starts to resemble the normal curve, and it fully conforms to normal distribution patterns when the sample size reaches 30. Specifically, smaller sample sizes, such as those below 20, can result in deviations from normal distribution patterns due to inadequate estimations of data dispersion. Therefore, for the subsequent analysis in this study, non-parametric methods were used for inference.

In Section 4.1, descriptive statistics are first presented and the repeated-measures ANOVA was utilized to identify between-group differences in overall accuracy rates. In Section 4.2, the Wilcoxon signed-rank test was employed to investigate differences within each group regarding their performance between the immediate and delayed tests. To examine between-group differences in the respective immediate and delayed tests, Section 4.3 employed the Mann-Whitney U test. In Section 4.4, the Friedman test was utilized to assess within-group differences in terms of accuracy rates across the three different contexts of T3 occurrence in both the immediate and delayed tests.

## 4.1 Descriptive Statistics of Overall Accuracy Rates

All 32 participants (16 in the CG were taught using the FT3F method, and 16 in the EG were taught using the HT3F method) completed both the immediate and delayed tests. This analysis excluded three individuals who did not pass the prerequisite test.

In the CG (n=16), the average accuracy rates in the immediate test were 53% for pre-T3, 93% for phrase-final T3, and 40% for T3 preceding T1, T2, or T4, resulting in an overall average of 62%. The average scores in the delayed test were 47% for pre-T3, 96% for phrase-final T3, and 43% for T3 preceding T1, T2, or T4, with an overall average of 62%.

In the EG (n=16), immediate test accuracy rates averaged at 89% for pre-T3, 91% for phrase-final T3, and 90% for T3 preceding T1, T2, or T4, resulting in an overall average of 90%. For the delayed test, the average scores were 64% for pre-T3, 85% for phrase-final T3, and 95% for T3 preceding T1, T2, or T4, with an overall average of 81%.

When considering both groups together (n=32), immediate test accuracy rates averaged at 71% for pre-T3, 92% for phrase-final T3, and 65% for T3 preceding T1, T2, or T4, resulting in an overall average of 76%. In the delayed test, the average scores were 55% for pre-T3, 91% for phrase-final T3, and 69% for T3 preceding T1, T2, or T4, with an overall average of 72%.

The repeated-measures ANOVA indicates that the EG exhibits higher overall accuracy rates compared to the CG when pronouncing T3 in disyllabic contexts. As shown in Table 5, the between-group effect of the *group* variable is statistically significant (F = 5.745, $p < .05$) concerning the overall accuracy rate. Post hoc comparisons reveal that the EG has a higher accuracy rate than the CG.

Table 5: Differences between the Groups' Overall Accuracy Rates (Repeated-Measures ANOVA)

| Source of variance | Sum of squares (SS) | Degree of freedom (df) | Mean square (MS) | *F* | *p* | Post hoc | Mean difference |
|---|---|---|---|---|---|---|---|
| Between-group effect | | | | | | | |
| Group | .389 | 1 | .389 | 5.745* | .023 | EG > CG | .185 |

* = p < 0.05

### 4.2. Within-Group Difference Analysis Regarding Retention

The results of the Wilcoxon signed-rank test, which was used to examine within-group differences in each group's performance between the immediate and delayed tests, are presented as follows:

For the CG, there is no significant difference in the accuracy rates of the tokens related to A (W = 0.954, $p > .05$), B (W = 1.809, $p > .05$), or C (W = 0.689, $p > .05$), as well as for the overall accuracy rates (W = 0.079, $p > .05$), between the immediate and the delayed tests. In contrast, for the EG, the accuracy rates in the immediate and delayed tests significantly differ for tokens associated with A (W = 2.101, $p < .05$). However, no significant differences are found for tokens associated with B (W = 1.172, $p > .05$), or C (W = 1.380, $p > .05$), as well as for the overall accuracy rates (W = 1.279, $p > .05$).

### 4.3. Between-Group Difference Analysis

Table 6 illustrates the comparison of accuracy rates between the EG and the

CG in both the immediate and delayed tests for the tokens associated with A, B, and C, as well as for their overall accuracy rates. The results indicate the following:

In the immediate test, the EG achieves significantly higher accuracy rates than the CG for the tokens associated with A and C, as well as for the overall accuracy rates. Moreover, in the delayed test, the EG also demonstrates higher accuracy rates for the tokens associated to C compared to the CG. Conversely, the CG exhibits significantly higher accuracy rates in the delayed test for the tokens associated with B when compared to the EG.

Table 6: Differences between the Groups' Accuracy Rates (Mann-Whitney U test)

| Accuracy rate of each environment of occurrence | | Group | Number | Mean rank | Sum of ranks | *U* statistic | *p* |
|---|---|---|---|---|---|---|---|
| Immediate test | A | CG | 16 | 12.00 | 192.00 | 56* | .004 |
| | | EG | 16 | 21.00 | 336.00 | | |
| | | Total | 32 | | | | |
| | B | CG | 16 | 16.53 | 264.50 | 127.5 | .984 |
| | | EG | 16 | 16.47 | 263.50 | | |
| | | Total | 32 | | | | |
| | C | CG | 16 | 11.38 | 182.00 | 46* | .001 |
| | | EG | 16 | 21.63 | 346.00 | | |
| | | Total | 32 | | | | |
| | Overall accuracy rate | CG | 16 | 10.81 | 173.00 | 37* | .001 |
| | | EG | 16 | 22.19 | 355.00 | | |
| | | Total | 32 | | | | |
| Delayed test | A | CG | 16 | 15.06 | 241.00 | 105 | .361 |
| | | EG | 16 | 17.94 | 287.00 | | |
| | | Total | 32 | | | | |
| | B | CG | 16 | 20.59 | 329.50 | 62.5* | .010 |
| | | EG | 16 | 12.41 | 198.50 | | |
| | | Total | 32 | | | | |
| | C | CG | 16 | 11.06 | 177.00 | 41* | .000 |
| | | EG | 16 | 21.94 | 351.00 | | |
| | | Total | 32 | | | | |
| | Overall accuracy rate | CG | 16 | 13.25 | 212.00 | 76 | .051 |
| | | EG | 16 | 19.75 | 316.00 | | |
| | | Total | 32 | | | | |

* = $p < 0.05$

## 4.4 Within-Group Difference Analysis Regarding Environments of Occurrence

The results from the analyses in Section 4.1 show that, overall, the EG achieves higher accuracy rates compared to the CG. This section provides a more detailed analysis to explore how the accuracy rates of the two groups in the three contexts of occurrence differ both in the immediate and the delayed tests.

As presented in Table 7, the accuracy rates of the EG for the items associated with the three environments of occurrence are not significantly different in the immediate test ($F = 0.176$, $p > .05$). However, in the delayed test, a noticeable difference becomes apparent ($F = 10.863$, $p < .05$), with the EG displaying higher accuracy rates for tokens associated with C compared to those associated with A.

Table 7: Difference Analysis of EG's Accuracy Rates of the Three Environments of Occurrence (Friedman Test)

| Time | Accuracy rate of each environment of occurrence | Mean rank | F statistic | *p* | Post hoc |
|---|---|---|---|---|---|
| Immediate test | Immediate test: accuracy rate of A (1) | 2.06 | .176 | .916 | |
| | Immediate test: accuracy rate of B (2) | 1.97 | | | |
| | Immediate test: accuracy rate of C (3) | 1.97 | | | |
| Delayed test | Delayed test: accuracy rate of A (1) | 1.63 | 10.863* | .004 | |
| | Delayed test: accuracy rate of B (2) | 1.78 | | | |
| | Delayed test: accuracy rate of C (3) | 2.59 | | | 3>1 |

* = $p < 0.05$

Regarding the CG, the Friedman test results, as shown in Table 8, indicate that the group's accuracy rates for the items associated with the three environments of occurrence are significantly different both in the immediate test ($F = 15.164$, $p < .05$) and in the delayed test ($F = 12.667$, $p < .05$). The CG consistently displays a higher accuracy rate for tokens associated with B than for those associated with A and C in both the immediate and delayed tests.

Table 8: Results of Difference Analysis of CG's Accuracy Rates of Three Environments of Occurrence (Friedman Test)

| Time | Accuracy rate of each environment of occurrence | Mean rank | F statistic | *p* | Post hoc |
|---|---|---|---|---|---|
| Immediate test | Immediate test: accuracy rate of A (1) | 1.78 | 15.164* | .001 | 2>1、3 |
| | Immediate test: accuracy rate of B (2) | 2.72 | | | |
| | Immediate test: accuracy rate of C (3) | 1.50 | | | |
| Delayed test | Delayed test: accuracy rate of A (1) | 1.75 | 12.667* | .002 | 2>1、3 |
| | Delayed test: accuracy rate of B (2) | 2.63 | | | |
| | Delayed test: accuracy rate of C (3) | 1.63 | | | |

* = $p < 0.05$

## 5. Discussion

### 5.1 Summary of Research Findings

This study investigates the effectiveness of the HT3F method for non-tonal speakers learning L2 Mandarin, comparing it with the FT3F method in terms of the accuracy of T3 pronunciation in disyllabic environments. The repeated-measures analysis shows that, in general, the EG achieves higher accuracy in pronouncing T3 in disyllabic contexts than the CG, as indicated in Table 5 of Section 4.1. This outcome aligns with the predictions based on the TMS and the CCH.

However, the statistical analyses reveal unexpected results, suggesting the influence of other significant variables in addition to the TMS and the CCH. These unexpected findings are listed below, indicating the complexity of factors affecting the results.

1. Within-group comparison reveals that the EG exhibits significantly higher accuracy rates of pre-T3 (token A) in the immediate test compared to the delayed test. Conversely, the CG does not demonstrate any significant difference between the two posttests in this regard.
2. Between-group comparison unexpectedly indicates that the EG does not

achieve a higher accuracy rate than the CG for phrase-final T3 (token B). In fact, the EG exhibits a significantly lower accuracy rate in the delayed test compared to the CG for this specific context.

## 5.2 Optimality Theory

The unexpected decline in the accuracy rate of pre-T3 by the EG and its unexpected underperformance compared to the CG on phrase-final T3 suggest the potential influence of other significant variables, possibly related to constraints within the framework of Optimality Theory (OT).

Within the OT framework, a general principle known as tonal faithfulness (McCarthy and Prince 1993; Prince and Smolensky 2004) plays a crucial role. Faithfulness constraints in OT require that "outputs preserve the properties of their basic (lexical) forms, requiring some kind of similarity between the output and the input" (Kager 1999). In the context of Mandarin T3 sandhi, the relevant faithfulness constraint is "MAX (contour)", which implies that correspondent segments in the input and output should have as many identical values as possible for tone contours.

For the pre-T3 sandhi rule, this faithfulness constraint, MAX (contour), provides a straightforward explanation for the decline in the accuracy rate of pre-T3 by the EG. This decline occurs because the HT3F method violates MAX (contour) through a categorical change from the base form [21] to the pre-T3 variant [35]. In contrast, the FT3F method, to some degree, adheres to this principle by involving a tonal reduction process from the base form [214] to the pre-T3 variant [14/35][5].

In accordance with Yin's perspective (2017:30) within the framework of OT, constraints are indeed violable, but any violations should be kept to a minimum. In the case of the FT3F method, it only partially violates the MAX (contour) constraint when implementing the pre-T3 sandhi rule. Conversely, the HT3F

[5] In accordance with Yin (2017) and Zhang (2018), this study utilizes the contour number [14] to represent the underlying representation of the rising portion of T3, even though the actual surface realization is [35]. Regardless of whether it is represented as [14] or [35], the fundamental characteristics of the rising portion of T3 is preserved.

method entirely violates this constrain, primarily because the pre-T3 sandhi rule involves a categorical phonological change from the base form [21] to the pre-T3 variant [35]. Consequently, it is suspected that this violation plays a significant role in the noticeable decrease in correct pre-T3 production by learners taught with the HT3F method.

Now, let’s delve into the unexpected underperformance of the EG compared to the CG on phrase-final T3 in the delayed test. Upon closer examination, individual analyses revealed that the EG made most of its phrase-final T3 errors in the delayed test when two T3s were adjacent. Specifically, in the delayed test, the accuracy rate of phrase-final T3 in T3-T3 combination was only 0.7 (SD = 0.24) for the EG, whereas the accuracy rate of phrase-final T3 in T1/T2/T4-T3 combinations was as high as 0.91 (SD = 0.13). In contrast, the CG did not exhibit such a discrepancy in the production of phrase-final T3. In the delayed test, the accuracy rate of phrase-final T3 in T3-T3 combinations was 0.94 (SD = 0.15) for the CG, whereas the accuracy rate of phrase-final T3 in T1/T2/T4-T3 combinations was 0.97 (SD = 0.11). In other words, despite being the same variant, half-T3 [21] was pronounced with the least accuracy by the EG only in T3-T3 combinations. It is speculated that the effect of violating the MAX (contour) constraint in the T3-T3 combination, as discussed earlier, might have spilled over to the second syllable in this combination and simultaneously led to a significantly lower accuracy rate of the second T3.

As discussed earlier, the violation of the MAX (contour) constraint by the EG has been suggested as a possible reason for the decline in the accuracy of pre-T3 and the poor performance on the second syllable in the T3-T3 disyllabic combination. However, it is important to note that this is a speculative explanation, and further studies are needed to verify and explore the exact mechanisms at play in these specific phonological phenomena. Additional research may provide a more comprehensive understanding of the interactions between phonological constraints and tonal production in the context of Mandarin T3 sandhi.

## 5.3 The CCH and the TMS in L2 Acquisition of Mandarin T3

The study's utilization of the DCM proposed by Jakubowicz (2011) to quantify complexity based on the number of steps required to acquire specific T3 variants in each teaching method is a valuable approach. The findings from this study generally support the notion that higher complexity, as measured by the number of steps required, can indeed increase processing burden and lead to lower accuracy.

In this case, the FT3F method, which introduces three T3 variants and thus requires learners to memorize and apply two sandhi rules: the pre-T3 sandhi rule and the half-T3 sandhi rule, is quantitatively more complex. According to the results presented in this study, the lower accuracy observed in the CG is likely the result of the extra processing required to apply the half-T3 rule. This sandhi rule, which is applied more frequently than the pre-T3 sandhi, would cease to exist if we instead assumed that half-T3 is the base form. In essence, the HT3F method is shown to be a less complex and more effective approach for learners to acquire T3 in disyllabic settings compared to the FT3F method.

The TMS also aligns with the findings of this study by predicting that the HT3F method should be a more effective way to teach Mandarin T3. According to the TMS, the ideal approach to T3 learning should start with the least marked variant before introducing the more marked ones. In the context of T3, the markedness ranking from high to low is full-T3 > pre-T3 > half-T3. However, the FT3F method takes the opposite approach by introducing T3 variants from the most marked to the least marked. This study's results support the TMS by indicating that the HT3F method, which aligns with the predicted sequence of markedness, leads to higher accuracy in T3 production.

It is important to note that the current study does not provide clarity on whether the outperformance of the EG is a result of the teaching sequence or the absence of full-T3 in instruction, as the HT3F method includes both of these factors in comparison to the FT3F method. Unfortunately, the data from our study do not allow us to disentangle these two compounded factors. Our assumption here is that the EG's superior performance is likely due to the combined influence of these two variables. However, it is crucial to test this assumption through a specifically designed experiment in our future research.

## 6. Conclusion and Pedagogical Implications

The present study fills a significant gap by conducting an experimental examination of the HT3F method in comparison with the traditional FT3F method. By doing so, it provides empirical evidence that substantiates the Tonal Markedness Hypothesis and the Computational Complexity Hypothesis, highlighting the importance of complexity and markedness factors in phonological acquisition. Furthermore, this study offers compelling evidence for the efficacy of the HT3F method in facilitating the acquisition of Mandarin T3 within disyllabic settings. It not only supports existing theoretical frameworks but also demonstrates their practical implications in language pedagogy.

The findings of this study hold particular significance, especially given the omission of half-T3 in prominent Mandarin Chinese textbooks used in Taiwan. It is hoped that these research results will serve as a compelling rationale for educators to consider and adopt the HT3F method in the instruction of T3 allophones. By doing so, educators can provide a more comprehensive and effective approach to teaching Mandarin T3, addressing an important aspect of the language that has been overlooked or inadequately covered in traditional teaching materials.

While the HT3F method has proven effective for T3 acquisition in disyllabic settings, it’s important to acknowledge that it may not fully account for all variables, as evidenced by the T3-T3 combination results. The categorical change introduced by the pre-T3 sandhi rule in the HT3F method can lead to constraint violations, impacting both pre-T3 and the second T3. To address this issue, instructors using the HT3F method should consider providing additional practice specifically targeting the T3-T3 combination to enhance retention and accuracy in this context.

## References

Broselow, Ellen, Su-I. Chen, and Chilin Wang. 1998. The emergence of the unmarked in second language phonology. *Studies in Second Language Acquisition* 20. 2: 261-280.

Chao, Yuen Ren. 1930. A system of tone letters. *Le Maître Phonétique* 45:

24-27.

Chen, Chung-yu, and Chilin Shih. 2021. The acquisition of Mandarin by heritage speakers and second language learners: the case of tone sandhi perception. *Journal of Chinese Language Teaching* 18. 4: 19-66.

Chen, Qinghai. 1997. Toward a sequential approach for tonal error analysis. *Journal of the Chinese Language Teachers Association* 32. 1: 21-39.

Chen, Yiya, and Yi Xu. 2006. Production of weak elements in speech–evidence from $f_0$ patterns of neutral tone in Standard Chinese. *Phonetica* 63. 1: 47-75.

Cheng, Chin-Chuan. 1973. *A Synchronic Phonology of Mandarin Chinese*. The Hague: Mouton de Gruyter.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris Publications.

Duanmu, San. 2000. *The phonology of standard Chinese*. Oxford: Oxford University Press.

He, Yunjuan. 2014. The factors affecting the production of Mandarin Chinese coarticulated tones by American learners. *International Chinese Language Education* 1: 136-146.

Hyman, Larry M., and Kenneth VanBik. 2004. Directional rule application and output problems in Hakha Lai Tone. *Language and Linguistics* 5. 4: 821-861.

Hyman, Larry M. 2007. Universals of tone rules: 30 years later. *Tones and Tunes: Studies in Word and Sentence Prosody*, ed by Tomas Riad and Carlos Gussenhoven, 1-34. Berlin: Mouton de Gruyter.

Jakubowicz, Celia. 2011. Measuring derivational complexity: new evidence from typically developing and SLI learners of L1 French. *Lingua* 121. 3: 339-351.

Kager, René. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.

Krithikadatta, Jogikalmat. (2014). Normal Distribution. *Journal of Conservative Dentistry*, 17. 1: 96-97

Li, Charles N., and Sandra A. Thompson. 1977. The acquisition of tone in

Mandarin-speaking children. *Journal of Child Language*. 4. 2: 185-99.

Major, Roy C. 2001. *Foreign Accent: The Ontogeny and Phylogeny of Second Language Phonology*. Mahwah, NJ: Lawrence Erlbaum.

McCarthy, John. J, and Alan Prince. 1993. *Prosodic Morphology: Constraint Interaction and Satisfaction*. Ms. Amherst & Rutgers: University of Massachusetts.

McCarthy, John. J, and Alan Prince. 1995. Faithfulness and reduplicative identity. *Papers in Optimality Theory*, ed. by Jill Beckman, Laura W. Kickey, and Suzanne Urbanczyk, 249-384. Amherst: Graduate Linguistic Student Association.

Mei, Tsu-lin. 1977. Tones and tone sandhi in 16th Century Mandarin. *Journal of Chinese Linguistics* 5. 2: 237-260.

Miracle, W. Charles. 1989. Tone production of American students of Chinese: A preliminary acoustic study. *Journal of the Chinese Language Teachers Association* 24: 49-65.

Ohala, John J. 1978. Production of tone. *Tone: A Linguistic Survey*, ed. by Victoria A. Fromkin, 3-39. New York: Academic Press.

Prévost, Phillipe, Nelleke Strik, and Laurie Tuller. 2014. *Wh*-questions in child L2 French: derivational complexity and its interactions with L1 properties, length of exposure, age of exposure, and the input. *Second Language Research* 30. 2: 225-250.

Prince, Alan, and Paul Smolensky. 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden & Oxford: Blackwell.

Shen, Susan X-N. 1989. Toward a register approach in teaching Mandarin tones. *Journal of the Chinese Language Teachers Association* 24. 3: 27-47.

Shi, 2007. On teaching tone three in Mandarin. *Journal of the Chinese Language Teachers Association* 42. 2: 1-10.

Tsung, Chin. 1987. Half-Third First: On the nature of the third tone. *Journal of the Chinese Language Teachers Association* 22. 1: 87-101.

Wang, Yue, Allard Jongman, and Joan A. Sereno. 2003. Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America* 113: 1033-1044.

Winke, Paula. 2007. Turning into tones: The effect of L1 background on L2 Chinese learners' tonal production. *Journal of the Chinese Language Teachers Association* 42. 3: 21-55.

Xu, Yi. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25: 61-83.

Yang, Chunsheng. 2016. *The Acquisition of L2 Mandarin Prosody: From Experimental Studies to Pedagogical Practice*. Amsterdam: John Benjamins.

Yin, Hui. 2017. An optimality theory perspective on Mandarin tone 3 sandhi. *Annals of Language and Literature* 1. 2: 28-34.

Yip, Moira. 2002. *Tone*. Cambridge: Cambridge University Press.

Yuan, Boping. 2015. The effect of computational complexity on L1 transfer: evidence from L2 Chinese attitude-bearing wh-questions. *Lingua* 167: 1-18.

Zhang, Hang. 2014. The third tone: Allophones, sandhi rules, and pedagogy. *Journal of the Chinese Language Teachers Association* 49. 1:117-145.

Zhang, Hang. 2018. *Second Language Acquisition of Mandarin Chinese Tones.* Utrecht Studies in Language and Communication 33 [Bilingual ed.].

Zhu, Hua, and Barbara Dodd. 2000. The phonological acquisition of Putonghua (Modern Standard Chinese). *Journal of Child Language* 27: 3-42.

國立臺灣師範大學國語教學中心(主編)。1999。《實用視聽華語》。臺北:正中書局。[National Taiwan Normal University Mandarin Training Center. (ed.) 1999. *Practical Audio-Visual Chinese*. Taipei: Cheng Chung Book Co., Ltd.]

國立臺灣師範大學國語教學中心(主編)。2021。《當代中文課程（第二版）》。新北：聯經出版公司。[National Taiwan Normal University Mandarin Training Center. (ed.) 2021. *A Course in Contemporary Chinese* (2nd ed.). New Taipei: Linking Publishing Company.]

[審查：2023.8.23 修改：2023.9.18 接受：2023.9.23]

黃韻芷
Yun-Chih HUANG
407224 臺中市西屯區臺灣大道 4 段 1727 號 東海大學
Tunghai University
No. 1727, Sec. 4, Taiwan Boulevard, Xitun District,
Taichung City 407224, Taiwan
yunchih.huang@gmail.com

黃郁欣
Yu-Hsin HUANG
407224 臺中市西屯區臺灣大道 4 段 1727 號 東海大學
Tunghai University
No. 1727, Sec. 4, Taiwan Boulevard, Xitun District,
Taichung City 407224, Taiwan
yuhsinhuang@thu.edu.tw

Appendix A: Change of the Third Tone in *Practical Audio-Visual Chinese* (National Taiwan Normal University Mandarin Training Center 1999)

**3. Change of the Third Tone**

**(1)**

ㄏㄣˇ ㄗㄤ hěn zāng　　ㄋㄧˇ ㄍㄠ nǐ gāo
ㄏㄣˇ ㄋㄢˊ hěn nán　　ㄋㄧˇ ㄇㄤˊ nǐ máng
ㄏㄣˇ ㄇㄢˋ hěn màn　　ㄋㄧˇ ㄌㄟˋ nǐ lèi

**(2) If two syllables in succession use the third tone, the first syllable changes to the second tone and the second syllable keeps the third tone.**

ㄏㄣˇ ㄏㄠˇ hěn hǎo → ㄏㄣˊ ㄏㄠˇ hén hǎo　　ㄋㄧˇ ㄏㄠˇ nǐ hǎo → ㄋㄧˊ ㄏㄠˇ ní hǎo
ㄏㄣˇ ㄌㄢˇ hěn lǎn → ㄏㄣˊ ㄌㄢˇ hén lǎn　　ㄋㄧˇ ㄌㄢˇ nǐ lǎn → ㄋㄧˊ ㄌㄢˇ ní lǎn

DRILLS 2 EVERYDAY LANGUAGE AND PRONUNCIATION

ㄏㄣˇ ㄗㄠˇ hěn zǎo → ㄏㄣˊ ㄗㄠˇ hén zǎo　　ㄋㄧˇ ㄗㄠˇ nǐ zǎo → ㄋㄧˊ ㄗㄠˇ ní zǎo

**(3) If more than two syllables of third tone are in succession, the tone changes according to context.**

ㄋㄧˇ ㄏㄣˇ ㄏㄠˇ nǐ hěn hǎo → ㄋㄧˊ ㄏㄣˊ ㄏㄠˇ ní hén hǎo / ㄋㄧˇ ㄏㄣˊ ㄏㄠˇ nǐ hén hǎo
ㄋㄧˇ ㄏㄣˇ ㄌㄢˇ nǐ hěn lǎn → ㄋㄧˊ ㄏㄣˊ ㄌㄢˇ ní hén lǎn / ㄋㄧˇ ㄏㄣˊ ㄌㄢˇ nǐ hén lǎn
ㄋㄧˇ ㄏㄣˇ ㄗㄠˇ nǐ hěn zǎo → ㄋㄧˊ ㄏㄣˊ ㄗㄠˇ ní hén zǎo / ㄋㄧˇ ㄏㄣˊ ㄗㄠˇ nǐ hén zǎo

**Appendix B: Change of the Third Tone in *A Course in Contemporary Chinese* (National Taiwan Normal University Mandarin Training Center 2021)**

2. Third Tone Change 三聲變調

(1) When two characters with third tones are found together, the first third tone is pronounced with a second tone as in these two examples in this lesson:

小姐 xiǎojiě（∨+∨→ˊ∨）　你好 nǐ hǎo（∨+∨→ˊ∨）

Despite the change in pronunciation, however, it is still marked as a third tone in pinyin.

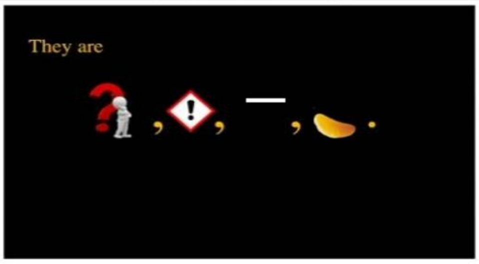
(2) Rules for third-tone changes 三聲變調原則：

| | | |
|---|---|---|
| ∨ ∨<br>(ˊ)<br>小 姐 | ∨ ∨ ∨<br>(ˊ)<br>我 很 好 | ∨ ∨ —<br>(ˊ)<br>很 好 喝 |
| ∨ ∨<br>(ˊ)<br>你 好 | ∨ ∨ ∨<br>(ˊ)<br>李 小 姐 | ∨ ∨ —<br>(ˊ)<br>我 喜 歡 |
| ∨ ∨<br>(ˊ)<br>很 好 | | ∨ ∨<br>(ˊ)<br>你 好 嗎 |

## Appendix C: Screenshots from the Treatment Video for the EG

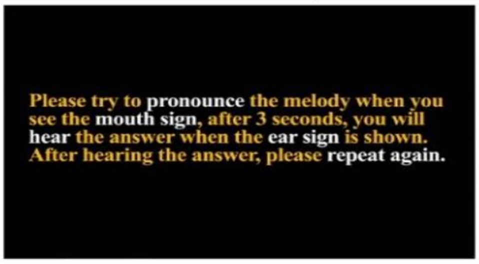
Thumbnail


Images represent four tones in the teaching order from left to right



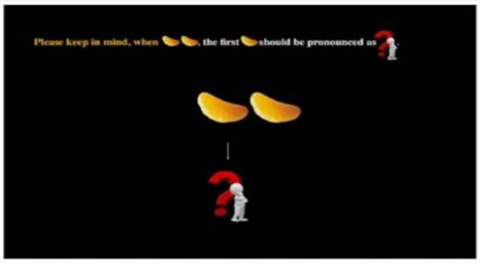

Explanation of the mouth symbol


Instruction for the recap session




Introduction of two T3 variants


Introduction of pre T3




## Appendix D: Targeted Stimuli Involving T3 in Posttests

| Environments of occurrence | Tokens | Numbers of tokens |
|---|---|---|
| Token A: Pre-T3 | [T3]+T3 | 4 |
| Token B: Phrase-final T3 | T1+[T3] | 4 |
| | T2+[T3] | 4 |
| | T3+[T3] | 0 (Identical tokens for pre-T3 are used.) |
| | T4+[T3] | 4 |
| Token C: T3 preceding T1, T2, or T4 | [T3]+T1 | 4 |
| | [T3]+T2 | 4 |
| | [T3]+T4 | 4 |

The total number of tokens for targeted stimuli involving T3 is 28.

# 華語 Half-T3 First 三聲教學之實證研究

**黃韻芷　　　　　　　　　　黃郁欣**
**東海大學華語文教學國際碩士學位學程**

## 摘要

本文研究華語聲調的教授和學習，並著重關注第三聲（T3）的教學議題。此議題是由對 T3 默認形式的不同理論假設發展而來的：傳統教學方法認為 full-T3（下沉輪廓，音高標示為[214]）是標準形式；而新創教法則以 half-T3（低頻，[21]）作為標準形式。由於目前沒有實驗研究驗證此兩種教學方法的效果，因此本研究進行了教學實驗研究來比較這兩種教學方法，並從中找出對第二語言學習者更有利的一種。本研究測試 32 位母語為無聲調的參與者，以兩階段的語言產出測試（立即後測和延遲後測）來評估參與者 T3 的表現。結果顯示，Half-T3 First（HT3F）的教學方法普遍提高了 L2 學習者在多數聲調組合中的三聲正確率。然而，在 T3-T3 的組合裡，研究參與者在第一音節三聲的正確率明顯消退，且第二音節三聲的表現不佳。此出乎意料的結果可歸因於優選理論（Optimality Theory）的忠實性原則（Principle of Faithfulness），而這則表示 HT3F 仍須考慮是否違反重要語音約束（Constraint）。

**關鍵詞**：Full-T3 First　Half-T3 First　計算複雜性假說　優選理論　聲調有標刻度