What is in the Final Stage of Inter-Language? Tone Errors and Phonological Constraints in Spontaneous Speech in Very Advanced Learners of Mandarin

Chenqing Song

Abstract Following the research direction proposed in Zhang (Journal of Chinese Language Teachers Association 45(1):39–65, 2010, The second language acquisition of Mandarin Chinese tones by English, Japanese and Korean speakers. [Doctoral Dissertation], University of North Carolina-Chapel Hill, 2013, Second language acquisition of Mandarin Chinese tones—Beyond first language transfer, Brill, 2018), this study examines the tone errors and substitution patterns in spontaneous connected speech produced by L2 learners who have progressed further into advanced or superior levels of proficiency. Based on the distinct patterns of errors and substitution patterns in their speech samples, effects of phonological universals in the format of constraints, including tone markedness scales (TMS), tone-position constraints (TPC) and the obligatory contour principle (OCP), are studied. Comparisons are made between the findings about tonal acquisition made in previous studies on lower-level learners and the higher-level learners in this study. With these error data, analyses and comparisons, I argue that some effects of the above-mentioned universals are still visible (TMS and T4-T4 OCP) while others are masked. It is a special configuration of the coarticulation rule applying to T2-T4 and T2-T1 combinations that really distinguishes the tonal system found in these very advanced learners from that of the other learners and that of the native speakers of Mandarin. Pedagogical practices that are designed to re-configure this rule will allow learners at this stage of tone acquisition to proceed into native-like speech production.

Keywords Phonological constraints · Tone errors in production · Advanced L2 learners

C. Song (✉)
SUNY-Binghamton University, Binghamton, NY, USA
e-mail: csong@binghamton.edu

1 Introduction

Mandarin tones have been studied from the aspect of phonology, phonetics, L1 and L2 acquisition. There are still unsolved problems, but generally researchers accept the following basic properties of the tones in Mandarin. There are four tone types for each full syllable and changing the tone will change the meaning of a morpheme, which is normally monosyllabic. The tones on single syllables may be transcribed and represented in different ways. Some of the most used ones are shown in Table 1.

The 1–5 scale was introduced in Chao (1930), in which 1 represents the lowest pitch point and 5 the highest. The HL system (following the tradition of Autosegmental Phonology, Goldsmith (1976)) uses H to represent a high tone, and L for a low tone and a contour tone is a combination of more than one-level tones. An M is sometimes used to indicate a mid-tone. The register-component tone representation is following the proposal in Bao (1999), in which register and contour of the tones are separate branches from the tone bearing unit (TBU), and the proposal made in Yip (1980), which uses [± upper] for the registers of tones. In this way of representation, h and l are terminal tone segments that are high and low, respectively. In this paper, tone category names and the HL system will be used to represent tones as they are already sufficient for the purpose of discussion here. Besides the four tones, Mandarin has a weak tone on reduced syllables, which is often called the "neutral tone." I am going to label it as Tn for the ease of recognition among other category labels.

When two or more tones are together, one or more of them may change the pitch value, a phenomenon called "tone sandhi." Mandarin's most widely discussed tone sandhi rule is the third tone sandhi (T3 sandhi), which states that a T3 will be pronounced as a high rising (MH) when it is followed by another T3. The resulting high rising is believed to be identical to the citation form of T2 but some researchers argue that the two are still distinctive (Zee 1980; Kratochivil 1987; among others). In this paper, I will label the output of the T3 sandhi as T5. When preceded by tones of a different category, T3 is pronounced as a low, slightly falling tone (ML), which is going to be labeled as HT3. This process is known as the half T3 sandhi. There are a couple of other sandhi rules, including some that are lexically restricted and a T2 Sandhi (Chao 1968) which occurs to T2s in some trisyllabic expressions. The domain of sandhi rules is a topic of prolific research, but it is not going to be touched upon in this paper.

Table 1 Basic information of the four tones in Mandarin

Category label	1–5 scale	HL system	Register + component tones	Example
T1	55	H	[+ U, hh]	mā "mother"
T2	35	MH	[+ U, lh]	má "hemp"
T3	214/21	L	[−U, ll]	mǎ "horse"
T4	51	HL	[+ U, hl]	mà "to scold"

Studies on the learning of Mandarin tones among L2 learners have produced rich insights into various issues. By studying the tone production and errors, researchers tried to identify the sources of the errors. Many of them (White 1981; Shen 1989; Miracle 1989; Guo 1993; Chen 1997; Sun 1998; Q.H. Chen 2000) focus on L1 transfer effects or the impact of L1 on the mastering of Mandarin tones. The complexity of the Mandarin tonal system and particular tonal features are also found to cause errors (Shen 1989; Miracle 1989; Elliot 1991, Hao 2012; among others). Some studies (Leather 1990; Elliot 1991; Guo 1993; Wang et al. 1999; Wang 2006; Hao 2012; Yang 2015; among others) combine production and perception, arguing that the relationship between the two plays a role in the accurate/inaccurate production of tones. In terms of research design, there are experimental studies, in which subjects are recruited to perform production or perception tasks with target tones[1] and combinations being solicited. Only a few studies are longitudinal and checked learners' performance at different developmental stages (e.g., Guo and Tao 2008). Yang (2011) investigates tone errors in the greater context of Mandarin phonology and argues that the tone errors are the results of "superimposition of the L1 English utterance-level prosody over tone production by L2 learners." Among the above-mentioned research, most are done on English native speakers. Zhang (2010, 2013, 2018) diverges from the other studies on the topic by focusing on the universal phonological constraints under the framework of the Optimality Theory (OT) (Prince and Smolensky 1993; McCarthy and Prince 1993, 1995; among others). By comparing the experimental data from three different learner groups (English native speakers, Korean native speakers and Japanese native speakers), L1 transfer effects are isolated from the possible effects of universal constraints. There are issues remaining in the study of the topic. As Zhang (2013) points out, research that bridges general linguistic theories and language teaching is still scarce. The data used for research are often isolated words, phrases and sentences with little discourse contexts. In experimental settings, tasks are often reading and repeating, which cannot reflect the proficiency of learners, an ability that can only be properly assessed in natural speech. Q.H. Chen (2000) tries to avoid this problem by using natural connected speech but found that tones in connected speech are hard to judge for accuracy out of context because tones (even produced by native speakers) in connected speech may be drastically different from their citation forms. Moreover, most existing literature on the L2 tonal acquisition is done on learners of low or intermediate proficiency levels, due to the limited availability of advanced-level or superior-level speakers in the past years.

---

[1]Target tone refers to the intended tone category that a speaker is trying to pronounce. In some studies, the target tones were provided to the speakers who read words/expressions from a list. In the current study, it is identified in the sample speech according to the lexical and contextual information. In connected spontaneous speech when a L2 learner's proficiency is low, and pronunciation errors as well as other errors are abundant, it is often difficult to identify the target tones. In the current study, the speech samples from the four speakers are very clear with almost no grammatical errors, lexical errors or segmental errors. Therefore, it is not difficult to know which word a speaker intended to say even if the tones were not pronounced accurately. Through the identification of words in the speech samples, the target tones were inferred.

The literature survey reveals gaps left unfilled by the previous studies and thus calls for a study in which spontaneous connected speech production of advanced or superior-level learners will be collected, examined and analyzed at the level of phonological acquisition. This is the approach and goal of this study. I will partially follow the framework proposed in Zhang (2010, 2013, 2018), especially focusing on three types of constraints: TMS, TPC and OCP. Tonal markedness scale (TMS) is a universal and phonetically grounded constraint, which ranks some tones as more marked than others. The proposed ranking or scale of the tones according to their markedness is *Rising ≫ *Falling≫*Level (Ohala 1978; Hyman and VanBik 2004). To L2 phonology, a more marked tone is often acquired later than a less marked one. Tone position constraints (TPC) are a set of constraints that states that certain tones are more marked or disfavored in certain positions. For example, Zhang (2004) proposes that "phrase-final syllables and syllables in shorter words are the preferred bearers of contour tones, even though they are usually not privileged for other phonological contrasts." Obligatory contour principle (OCP) is a constraint that states certain consecutive identical segments/tones/features are banned or disfavored (Goldsmith 1976).

Through controlled experiments done with three L1 groups (English, Korean and Japanese native speakers) and comparisons among the data obtained from the three groups, Zhang was able to make specific claims about the L1 transfer effects. Since only English native speakers' speech is used in this study, no cross-linguistics or L1 transfer claims will be made. Due to the length restriction of this paper, OT theory (and the corresponding OT L2 theory) will be brought into discussion at the conceptual level where rankings of relevant constraints are proposed but not at the technical level where all constrains are ranked in tableaus to output the attested forms.

2   Research Questions, Method and Subjects

Advanced-level (or superior-level) L2 learners of Mandarin Chinese are often not abundant, which explains why most previous studies focused on learners who are at lower levels of proficiency. In experiment-based studies, the sample size requirement compels researchers to choose subjects of study who are beginner or intermediate level learners. Not many studies have revealed the tonal productions among advanced or superior-level learners of Mandarin Chinese. As a consequence, many questions remain unanswered with regard to the tonal production of this group. To that end, the current study attempts to address some of these questions:

1. What kind of tone errors do advanced/superior learners make?
2. Do the errors they make fall into some common categories? If so, what are the categories of these errors?
3. Are these errors similar to or different from the errors identified from previous studies done mostly on learners of lower proficiency levels?

4. If so, can phonological universals (TMS, TPC and OCP) explain these errors? To what extent, do they explain these errors?

5. What do these errors reveal to us about the development of the inter-language (IL) at this very final stage of acquisition?

The significance of the answers to these questions extends beyond the understanding of Mandarin tonal acquisition. If the learners are at a very high level of proficiency, then their IL systems are very close to that of the native speakers. Then, the differences between the IL systems and the native speakers' system should provide clues about the Mandarin phonological system itself. As previously stated, the present study follows the theoretical standpoint in Zhang (2010, 2013, 2018), which views IL as an L2 phonological system, in which phonological constraints interact in different rankings to produce the observed outputs (see the OT framework mentioned above). Along this line of thinking, the crucial question to ask in addition to the five listed above is what differences exist among constraints and their rankings, which set apart the IL system and the native speakers' system. If the OT theoretical framework is adequate for L2 acquisition study, and our analysis is accurate, then it is predicted that there will only be a few constraints and minor differences in their rankings between the advanced learners and the native speakers. The results from the current study, to be presented in this paper, provide support to this hypothesis.

To answer these research questions, this paper uses spontaneous connected speech from four L2 speakers that are publicly available through YouTube channels. These speech samples were produced not for educational or research purposes but purely for informative and recreational ones. In other words, the speakers focused on the message over the forms. These speakers' pronunciation of the Mandarin tones is very close to that of the native speakers' with an error rate between 1.5% and 6.3%.[2] They are able to sustain a speech in Chinese for an extended period of time, as demonstrated by their YouTube videos, on a good variety of topics from education to politics, from daily life to economy. Their speech in Chinese is both fluent and accurate, facilitated by an extended vocabulary and a solid mastery of grammar. There is almost no grammatical error or misuse of lexical items in their speech. In a single video continuous shot with minimal post-shooting editing or revision, these speakers are able to elaborate on a topic for over ten minutes, weaving discourses that are highly consistent, coherent, culturally appropriate and functionally adequate. Using the ACTFL *Performance Descriptors for Language Learners* (ACTFL 2012) to evaluate these speakers, in the presentational mode of communication, two ACTFL OPI trained evaluators (including the author of the paper) independently assessed and confirmed that they fall into the advanced range (or higher) properly by demonstrating the following performance.

---

[2]The term "error rate" is defined as the ratio of total number of incorrect tones in all syllables in the sample for each speaker. The method of how the accuracy of tonal production was assessed will be explained in a later section of this paper. Initially, nine speakers were included in the pool. Five of them turned out to have much higher overall tone error rates and were assessed as lower in proficiency level using the ACTFL Performance Descriptor. So these five were elimited from the research subject pool. The four remaining speakers included in this study are those whose tone error rate is smaller than 10% and whose performance overall is at or above advanced level.

Descriptions of the speech of the subjects:

- Functions: They all produced narrations and descriptions on both familiar topics (such as learning a foreign language), and unfamiliar topics (such as Covid-19 pandemic). In their speech, they constructed well-supported arguments, including details of evidence in support of a point of view.
- Context/content: In their speech, they covered content areas that are of both personal and general interest. Also, there is some evidence showing that they are able to dive into more abstract notions (such as freedom or identity).
- Text type: They were producing paragraphs that are organized and detailed.
- Language control: Native speakers of Chinese with no training or experience working with nonnative speakers would have no difficulty understanding their speech. They master the grammar and syntax well with few mistakes. They use special constructions such as the Ba-construction or the resultatives accurately both in terms of forms and in terms of functions.
- Vocabulary: They used a good variety of vocabulary that are suitable for the topic and the contexts.
- Communication strategies: Although self-correction is not abundant (since they are already very accurate), there is clear evidence for elaboration, clarification and circumlocution.
- Cultural awareness: These speech samples are delivered in culturally appropriate manners and the speakers demonstrated cultural knowledge in their presentation of the topics.

Although evaluators are not able to conduct full OPI interviews and to test all the aspects of their speech performance to establish a performance ceiling, the floor of their proficiency is at the ACTFL advanced level. Without probing into a higher level, it is unknown where the ceiling of their proficiency is.

There are advantages and disadvantages using the spontaneous connected speech samples from publicly accessible platforms such as YouTube. The main advantage of using these spontaneous speech samples is that they reflect the natural status of the L2 language. These clips were produced for non-research purposes, with an intention that centers on the message (content) rather than linguistic forms, including the tones. The speakers each had a personal channel on YouTube, where they release such videos frequently covering topics from life to study, from society to politics. The samples were taken from their channels randomly. To ensure the consistency of performance, the selected samples are produced within a year of time for each speaker. The downside of using these spontaneous speech samples comes from a few directions. First, the quality of the sound is not ideal. The recording equipment, recording environment and the recording skills all affected the quality of the sound. These clips are definitely not research materials that can be used for acoustic analysis. Therefore, in the present study, trained native speakers' judgments are employed to assess the data and mark errors. This method is appropriate with regard to the data but may have missed important nuances in tonal production. The second issue with spontaneous connected speech concerns the representativeness of the data, an issue raised by many SLA researchers (see C. Chaudron 2003 for general discussions of

Table 2 Proportions of the four tones in each speakers' samples

	J (%)	M (%)	F (%)	X (%)
T1	15.9	17.1	17.0	17.4
T2	19.8	18.0	17.6	16.5
T3	20.7	19.2	18.6	20.9
T4	30.7	32.9	34.0	31.4
Tn	11.0	10.6	10.7	11.4

SLA data collections). It is well known that L2 speakers avoid the structures, words and sounds that they do not master well. In spontaneous speech, we may not observe what they do not do well, and studying what is present in the samples may not reveal the entire picture of their L2 status. In the present study, this issue is not very serious because each of the four tones are so abundant that it is impossible to completely avoid using a particular tone or a tonal combination. Moreover, the overall proportions of the four tones stay rather consistently across each speech clip and across speakers,[3] as shown in Table 2.

The third issue is not a problem but rather a disadvantage of spontaneous speech data collection compared to experimental data collection. In spontaneous speech, the forms/problems under investigation may not appear frequently enough, resulting in insufficient data for further analysis. To solve this problem, a large amount of spontaneous speech must be collected and analyzed in order to collect enough relevant data to answer the research questions, while in experimental studies, the target forms are designed to be solicited and produced by the subjects. In the present study, since the speakers are fluent and the length of their speech clips are long enough, there are enough tones of each category and each combination for the purpose of production error study. These speakers also produce nearly error-free speeches, so that transcribing and scoring errors are not as difficult as working on speeches from lower proficiency level L2 speakers of Chinese. As the data will show in the next section, not only is the overall number of errors in each speaker's tonal production small, but the types of errors are limited to a very small number of tone categories and tone combinations. Moreover, the error substitutions are almost all from the Mandarin tone inventory. The four speakers are all native speakers of English, but they learned Mandarin in different settings (e.g., different Chinese programs, teachers, etc.). The following table lists the basic information about these speakers (Table 3).

In Zhang (2010, 2013, 2018), subjects who speak different first languages form different groups in experimental settings. This allows Zhang to compare data from the different groups and isolate the L1 transfer effect from the influences from other factors including linguistic universals and pedagogical choices. In the present study, due to the limited availability of the speech samples from speakers at this proficiency

---

[3]There are not many well-reported studies on the proportions of tones in native speaker's natural speech. The estimates have been that T4 has the highest proportion, and L2 the lowest. FLA studies such as (Van de Wijer and Sloos 2014) used high frequency lexical words to calculate the proportion of the tones in L1 development and reached similar conclusions.

Table 3 Basic information about the four subjects[a]

Subject	Gender	Age	Native language	Studied and/or lived in China
J	M	20–25	English (Canadian)	10 years
M	M	31	English (American)	9 years
F	M	30	English (British)	10 years
X	F	26	English (American)	8 years

[a]This information was obtained from the speakers' public social media accounts such as YouTube channel, Bilibili.com channel or Weibo account

level, the four speakers included in this research all speak English as their native language. In future studies, more speech samples from speakers of other languages should be included to draw a more complete picture of the tone errors and error patterns.

The four speakers each have many videos on various topics on their YouTube channels. Clips (total length of 20 min or so) from each speaker were randomly selected[4] and transcribed into Chinese characters and pinyin by four experienced Chinese language teachers (8-10 years of teaching experience). Transcriptions were reviewed by at least one other transcriber. The speech of these four speakers is very close to that of the native speakers, so the transcribers found only one or two words in each of the video clips that they did not agree on or they could not decide what they heard. These words were marked with a few possible interpretations. Then two Chinese phonologists (including the author) listened to the video clips and assessed the tonal productions of each syllable and marked them as either "correct" or "incorrect," based solely on their language intuition in that connected speech context. In other words, the evaluations were not based on citation form or any standard form. If marked as "incorrect," the substitute tone was transcribed using the 1-5 scale tonal marking system (Chao 1930, 1948). Each of the phonologists did so independently and did so at least twice, with a minimum of one month between each transcription. The results from the two phonologists have an overall 98.3% rate of agreements. Where the two phonologists disagreed, a third phonologist was invited to make an independent judgment. A final discussion of the three phonologists resolved most disagreements except in two cases. These two cases were excluded from the data calculation. The following table shows the basic information of the video clips and the transcriptions (Table 4).

After tone errors were identified and transcribed, tonal contexts were also annotated for further analysis. Using these raw materials, here are the actions taken to answer the research questions raised above.

1. What kind of tone errors do these learners make?

    Action 1: Check the tonal errors found in our transcriptions. Identify the target tonal categories and the substitutions.

---

[4]In these speakers' channels, the lengths of the video clips are very different from each other. Very short and very long ones were excluded from the total before the random selection.

Table 4 Basic information of the speech samples

Speaker	Topics	Total length	Total syllables	Total incorrect tones
J	Cultural comparison Language learning Personal identity Personal history	21'42"	4670	117 (2.5%)
M	Covid-19 Language learning	22'36"	5694	88 (1.5%)
F	Covid-19 Brexit	24'17"	3884	106 (2.7%)
X	Covid-19 Language learning	18'45"	4733	299 (6.3%)

2. Do the errors they make fall into some common categoreis? If so, what are the categories of these errors?

   Action 2: Based on the results from step 1, compare data from different speakers.

3. Are these errors similar to or different from the errors identified from previous studies done mostly on learners of lower proficiency levels?

   Action 3: Based on the results from Step 2, compare the findings among the four speakers in this study to the findings in previous studies.

4. If so, do phonological universals (TMS, TPC and OCP) explain these errors? To what extent, do they explain these errors?

   Action 4: The results from Step 2 and 3 will be discussed to identify the causes of the errors, following part of the framework laid out in Zhang (2010, 2013), in which OPC, TMS and TPC are the three types of constraints among the universal phonological constraints.

5. What information do these errors reveal to us about the development of the IL at this very final stage of acquisition?

   Action 5: Rather than a separate step, it is more of a co-occurring action with Action 4. During the discussions about the causes of the errors, native speakers' tonal phonology will be brought up and used as a point of reference. By doing so, we will learn about the stage of development of IL and project the changes that are needed to further approximate the L1 tonal phonology of Mandarin Chinese. The universal phonological constraints and their ranking (and re-ranking) will be used as the theoretical tools in the discussion as well.

3 Data and Analysis

3.1 TMS

In the previous section, research questions and actions have been laid out. In this section, Actions 1 to 4 will be carried out. Action 5 will be done in Sect. 4 of the paper. Tables 5 and 6 show the errors in terms of tonal categories for each of the four speakers. In Table 5, the target tone categories are listed in the first row. In each speaker's row in each cell, the first number is the total number of errors in this target

Table 5 Tone errors in each tone category and their proportion in the total errors made by each speaker

	T1	T2	HT3	T4	FT3	T5	Tn	Total
J	18	80	6	13	0	0	0	117
	15.4%	68.4%	5.1%	**11.1%**	0%	0%	0%	100%
M	0	64	4	18	0	2	0	88
	0%	72.7%	4.5%	20.5%	0%	2.3%	0%	100%
F	12	73	3	15	1	1	1	106
	11.3%	68.9%	2.7%	14.2%	0.9%	0.9%	0.9%	100%
X	13	169	7	108	1	0	1	299
	4.3%	56.5%	2.4%	36.1%	0.3%	0%	0.3%	100%
Overall	43	386	20	154	2	3	2	610
	7.0%	63.3%	3.3%	25.2%	0.3%	0.5%	0.3%	100%

Table 6 Tone errors in each tone category and their proportion in the total number of targets

		T1	T2	HT3 + FT3	T4	T5	Tn	Total
J	Total target no. of errors	741	924	966	1434	93	512	4670
		18	80	6	13	0	0	117
	Percentage	2.4%	8.7%	0.6%	0.9%	0%	0%	2.5%
M	Total target no. of errors	975	1023	1094	1875	124	603	5694
		0	64	4	18	2	0	88
	Percentage	0%	6.3%	0.4%	1.0%	1.6%	0%	1.6%
F	Total target no. of errors	661	684	723	1319	83	414	3884
		12	73	4	15	1	1	106
	Percentage	1.8%	10.7%	0.6%	1.1%	1.2%	0.2%	2.7%
X	Total target no. of errors	825	781	988	1487	112	540	4733
		13	169	8	108	0	1	299
	percentage	1.6%	21.5%	0.8%	7.3%	0%	0.19%	6.3%
Overall	Total target no. of errors	3202	3412	3771	6115	412	2069	18981
		43	386	22	154	3	2	610
	Percentage	1.3%	11.3%	0.6%	2.5%	0.7%	0.1%	3.2%

Table 7 Tests comparing the error rates between tone categories

	T2 versus T4	T4 versus T1	T1 versus T3	(HT3 + FT3) versus T5
p value	**<0.0001**	**<0.0001**	**<0.001**	0.36
z	17.80	3.75	3.29	0.36

tone category and the percentage below this number is the proportion of this number to the total number of errors made by this speaker. For example, J made a total of 18 errors for target T1 and that 18 makes 15.4% of the total 117 tone errors in J's speech sample.

From Table 1 in the previous section, it is seen that each tone category has different numbers of total tone production in the speech. Some tone categories, such as T4, have higher frequency of occurrence than others. In natural native speaker's speech as well as the speech samples collected from these four speakers in this study, the number of tones in each category is not evenly distributed. In other words, some tones simply occur more frequently than others. One would expect that the more often a tone category is pronounced, the more tone errors in this category will occur. So higher frequency of tone errors in one tone category in a sample does not indicate that tone category is more difficult to pronounce. To avoid such possible misrepresentation, in Table 6, the numbers of tone errors are compared to the total target tones in each category, so that we can see whether there are more errors produced for each tone category proportionally.

The occurrences of errors given in the two tables above both show that the four speakers made more mistakes when they pronounced a T2. In three of the four speakers, T4 comes second. T3, whether in the HT3 subcategory, FT3 subcategory or in the T5 subcategory, has very few mistakes. When testing the numbers in Table 6 using statistical *z* tests, the following results, as shown in Table 7, are produced: the error rates of T2 are significantly higher than those of T4, which is significantly higher than those of T1, which is significantly higher than those of T3. Between the error rates found for T3 and T5, there is no significant difference.

The ratio of errors in each tone category and the statistical test results provide support to the discoveries found in previous studies (Miracle 1989; Shen 1989; Leather 1990; Elliot 1991; Guo 1993; Chen 1997; Sun 1998; Zhang 2010, 2013). In these studies, relative difficulty of the four tones to L2 learners is summarized in Table 8. The asterisks used in "the order of difficulties" indicate disfavor following the OT convention used in Zhang's studies, so *T2 > *T3 means that T2 is more difficult than T3.

In most previous studies, T1 has been shown to have a low error rate in production. There are different conclusions regarding the most difficult tone for L2 learners, possibly due to different kinds of evidence found in different production tasks (repeating, responding, reading or speaking). T2 appears to be difficult, ranking as the most or the second most difficult. Different reasons were given to explain the

Table 8 Ranking of difficultness from previous studies

Study	Order of difficultness	Length of subjects' experience of learning Chinese l
Miracle (1989)	*T2 > *T3 > *T4 > *T1	At least 1 year
Shen (1989)	T4* > *T1 > *T3 > *T2	4 months
Leather (1990)	*T3 = *T2 > *T4 > *T1	Minimal
Elliot (1991)	*T3 > *T2 > *T4 > *T1	1 semester to 400 level
Chen (1997)	*T3 > *T2 > *T4 > *T1	1–2 years
Sun (1998)	T2 being the most difficult	1–3 years
Guo and Tao (2008)	*T3 > *T2 > *T4 > *T1	1 and 2 semesters
Zhang (2010, 2013)	*T2 > *HT3 > *T4 > T5 > *T1 (English L1 speakers)	6–18 months

ranking of error rates in these studies, ranging from articulatory reasons, to production–perception relationship, to L1 prosodic transfer effects, and to pedagogical practices. In Zhang (2010, 2013, 2018), TMS was proposed to be the cause. According to the TMS *Rising ≫*Falling ≫*Level (Ohala 1978, Hyman and VanBik 2004) and *High≫ *Low (Yip 2002), Zhang proposes that in Mandarin, T2 is more difficult than T4, which is more difficult than T1 (Ranking *FT3 ≫*T2 ≫*T4≫*T1 ≫*HT3 (Zhang 2013)). The results from these studies have been used to answer the question regarding the "order of tone acquisition," assuming that the tones that are pronounced with a higher rate of errors are more difficult to acquire and will be acquired later. This assumption has not been verified by longitudinal studies. The results from our study provide valuable insight into this issue by revealing the actual tone production at the very end of the acquisition process, if there is such an order. The numbers in Table 5 show that T2 errors are more than half of the total errors in all four speakers. In fact, the proportion of T2 errors in the total number of errors increases when the speaker's overall tone error rate decreases as shown in Table 9. This may indicate that as speakers master tonal production better, errors in other categories disappear faster than those in T2, resulting in more T2 errors in the total number of errors. Unfortunately, the current study has only four speakers and this stipulation about developmental change deserves further study.

Next to T2, T4 has the second highest error rate in three out of the four speakers. In many previous studies (Leather 1990; Elliot 1991; Chen 1997; Guo and Tao 2008; Zhang 2010, 2013, 2018), T3 is reportedly difficult, while in our data, T3, including all three of its variants, has a very low error rate. Summarizing the results from

Table 9 Proportion of T2 error and the overall error rate of each speaker

	M	J	F	X
T2 error in all tone errors	**72.7%**	**68.4%**	**68.9%**	**56.5%**
Overall error rate (number of errors as a percentage of total syllables produced)	1.5%	2.5%	2.7%	6.3%

Tables 5 and 6, the rank of difficulty is *T2 >* T4 > *T1 > *T3 (including HT3, FT3 and T5). This is very close to what TMS predicates in Zhang (2013), except that FT3 is not problematic among the four speakers in the current study. Containing three tone targets, FT3 is a contour tone of the most complicated type. The fact that in all four speakers' speech, it is almost error free may be explained in a few ways. First, these four speakers are all very advanced. They have mastered the articulatory mechanisms of tones well. Secondly, in connected speech, FT3 is not abundant and has predictable occurrences, namely in isolated syllables and in prosodic-final positions. The fact that FT3 is almost problem free in these speech samples suggests that the causes for tone errors at this stage of IL are beyond articulatory reasons. In Zhang (2013), the high error rate of T3 found in the experiment (*T2 > *HT3 > *T4 > T5 > *T1 (English L1 speakers)) does not fit the predicted outcome based on the proposed TMS (*FT3 ≫ *T2 ≫ *T4 ≫ *T1 ≫ *HT ≫). The proposed explanation is the inappropriate pedagogical practice that emphasizes FT3 as the underlying form of T3. The outcome in the present study, in which HT3 behaves the way that the TMS predicts, suggests that the learners' IL may have gone through significant revisions, through which the underlying form (base form) of T3 was reset from FT3 to HT3. Another interesting point to make here regarding the errors in different single tone categories is about T5 or the sandhi form of T3 when it is followed by another T3. There are phonological and phonetic studies on the very nature of this tone particularly around the issue of whether it is distinguishable by native speakers or if it has distinguishable phonetic cues from those of real T2's (Zee 1980; Kratochvil 1987; Xu 1997 to name a few). In our data, T5 has a similar error rate to other variants of T3, which is much lower than that of T2. But we will see in the following discussion that this is related to the distributional properties of T5 and does not really provide many clues with regard to the debate surrounding T5's distinctiveness from T2 in Mandarin Chinese.

After examining the error rates in different single target tone categories, we now move on to the substitute forms of these errors. Previous studies of tone errors in L2 learners of Mandarin reveal that learners often produce tones that are out of the L1 Mandarin tone inventory (including both the citation forms and the legit variants in connected speech). Moreover, in connected speech, L2 speakers produce certain tonal variants (such as the "mid-tone" proposed in Q.H. Chen 2000) that are different from the citation form of the tones but are found in L1 speakers' connected speech. However unlike the L1 speakers, L2 speakers pronounced these variants in illegitimate locations. In the current research, tones in the connected spontaneous speech samples were judged by native speakers against the acceptable form(s) in native speakers' speech. If such variants occur in the places where native speakers would accept as possible forms, then they are not counted as errors. In other words, a tone is counted as an error only if it cannot be accepted by a trained native speaker in these contexts. Therefore, many "mid-tones" were counted as acceptable productions. In the speech samples in this study, the vast majority of the substitute forms do not fall outside of this inventory, and this is not unexpected for this group of high proficiency learners. Table 10 shows numbers of the erroneous substitute forms for the corresponding target tones. For example, in J's speech samples, he mispronounced T1 18

Table 10 Tone errors and their corresponding substitutes

	T1	T2	HT3	T4	FT3	T5	Total error
J	18	80	6	13	0	0	117
	HT3: 3 T4: 14 Mid-tone: 1	HT3: 79 T1: 1	T1: 4 T4: 2	T1: 13			
M	0	64	4	18	0	2	88
		HT3: 61 T1: 3	T1: 4	T1: 16 HT3: 2		HT3:2	
F	12	73	4	15	1	1	106
	HT3: 9 T4: 3	HT3: 66 T1: 7	T1: 3 T4: 1	T1: 12 HT3: 3	T1:1	HT3: 1	
X	13	169	9	108	1	0	299
	HT3: 9 T4: 4	HT3: 159 T1: 10	T1: 9	T1: 105 HT3: 2 Mid-tone: 1	T1: 1		

times, in which 3 of them were pronounced as HT3, 14 of them were pronounced as T4 and the other one was pronounced as a mid-tone.

The substitution patterns among the four speakers are strikingly similar. Overall, HT3 is the most frequent substitute for a mistakenly pronounced target tone (70% in J's speech, 74% in M's speech, 75% in F's speech and 57% in X's speech), followed by T1 and with T4 coming third. The substitute pattern confirms that the effect of TMS (*FT3 ≫*T2 ≫*T4 ≫*T1 ≫*HT3 (Zhang 2013)), which predicts that the low-level tone HT3 is the most unmarked member of the tonal system, and T1, being a high-level tone, is the second most unmarked. This substitute ranking generally aligns with the ranking found in Zhang (2013) among English speakers but there are a few minor differences between the two. First, HT3 is not only the most common substitute but also is the majority in the current study. In Zhang (2013), HT3 substitutes constitute only about 15% of the total. Secondly, although T2 and FT3 were found as substitute in Zhang (2013), in the current study, they do not appear in the speech as substitutes for a failed target tone production at all. Thirdly, T4 only appears 24 times as a substitute in our data out of the total 610 tone errors. Such differences between low/mid-level L2 speakers and advanced-level speakers suggest that the effect of TMS becomes more salient as learners' IL progresses and the effects from other factors, such as L1 negative transfer or pedagogical reasons, may have gradually faded away. When target tone information is added to the discussion, we find that: HT3 is the most frequent substitute for a target T1 (except in J's case where T4 is the most frequent substitute) or T2; T1 is the most frequent substitute for a target T4; T1 is the most frequent substitute for a target HT3. Table 11 compares our findings with the findings in Zhang (2013), where the differences are highlighted in gray.

Table 11 Ranking of difficultness in Zhang (2013) and in the current study

	Zhang (2013) (English speakers)	The current study
T1's substitute	HT3 > T4 > T2 > FT3	HT3 > T4
T2's substitute	HT3 > T1 > T4 > FT3	HT3 > T1
HT3's substitute	FT3 > T4 > T1 > T2	T1 > T4
T4's substitute	HT3/T1 > T2 > FT3	T1 > HT3
T5's substitute	HT3 > T1 > T4	HT3 (only a few cases total)

Compared to Zhang (2013)'s findings based on low-mid-level learners, the findings in this study show that the four sampled speakers have mastered the T3 sandhi with only 3 T5 errors. Secondly, in Zhang (2013) data, FT3 often shows up as the substitute of HT3, while in the current study, FT3 is not produced by any of the speakers in place of a target HT3. Instead, T1, the second most unmarked tone, appears in these places. Zhang (2013) attributes the HT3 substitute pattern to pedagogical practice. Along this line of reasoning, then the HT3 substitute pattern in our study suggests the effect of TMS surfaces more visibly among more advanced learners, who seem to have overcome the negative impact of the common pedagocial practice that, according to Zhang, promotes FT3 as the base tone. Thirdly, in Zhang (2013) data, HT3 is the most frequent substitute for all other tones (T1, T2, T4 and T5). This is almost the case in our data, except that we find the most frequent substitute for T4 is T1 by a large margin over the second most frequent substitute tone HT3. Therefore, TMS alone cannot explain the pattern here because it will predict that HT3 is also the most frequent substitute for T4. In the following section, examination of tone errors and substitution patterns in tonal combination contexts will provide a close-up view of the issues.

3.2 TPC

After examining the errors in single tones without taking into considerations its contexts, we now move onto tone error distributions within local contexts, mostly in prosodic words. Table 12 summarizes the tone errors and their position in words. In the table, "initial" means when a tone error occurs in the first syllable of a disyllabic (prosodic) word, "final" means an error occurs in the second syllable of a disyllabic (prosodic) word, "others" include mostly monosyllabic and polysyllabic words. For each speaker, under each tone category, the first row is the total number of tone errors in this tone category and then the cell below the total is broken down into number of tone errors in different positions.

HT3, FT3 and T5 have positional distributions via sandhi rules. So, the error distributions are affected by their allotonic distributions. Also, the errors in these three categories are few, so we will focus on the distributions of T1, T2 and T4 only. In all four speakers' speech, there are more errors found for a target T1 when it is at

Table 12 Tone errors in different positions in a prosodic word

	T1	T2	HT3	T4	FT3	T5	Total
J	Total 18	80	6	13	0	0	117
	Initial: 12 Final: 2 Others: 4	Initial: 61 Final: 11 others: 8		Initial: 10 Final: 1 Others: 2			
M	0	64	4	18	0	2	88
		Initial: 53 Final: 6 Others: 5		Initial: 8 Final: 9 Others: 1			
F	12	73	4	15	1	1	106
	Initial: 9 Final: 3 Others: 0	Initial: 56 Final: 13 Others: 4		Initial: 7 Final: 6 Others: 2			
X	13	169	8	108	1	0	299
	Initial: 11 Final: 1 Others: 1	Initial: 75 Final: 71 Others: 23		Initial: 40 Final: 54 Others: 14			
Total	43	386	23	154	2	3	610
	Initial: 32 Final: 6 Others: 5	Initial: 245 Final: 101 Others: 40		Initial: 66 Final: 69 Others: 19			

word-initial positions. When T2 is the target pronunciation, more errors are produced when it is word -initial in three of the four speakers individually and overall. T4 is different in this sense from T1 and T2 because overall the number of errors at initial versus final positions is roughly equal. There are individual differences among the four speakers, especially in the case of X, who also made more errors in speech than the other three speakers. In Zhang (2013), the TPC were investigated by comparing error rates (percentage of errors at either word-initial or word-final positions) for each tone category. Table 13 is the results for the English speakers in Zhang's study.

Zhang (2013) argues that the data in the table above (and the corresponding substitution patterns) demonstrate that "T2 is performed better at word-initial positions,

Table 13 Error rates with regard to positions in a word for English speakers (adapted from Zhang 2013: 111)

	General		English speakers	
	Initial (%)	Finall (%)	Initiall (%)	Finall (%)
T1	18.02	25.42	24.38	30.00
T2	33.33	79.9	36.88	78.13
T4	44.17	22.29	54.69	28.44

Table 14 TPC rankings

*Contour-I ≫ Contour-F (Word-final positions are better bearers of contour tones)
*Fall-I ≫*Fall-F (Falling tones are more disfavored at word-initial positions than at final positions)
*Rise-F ≫ *Fall-F (Rising tones are more disfavored than falling tones at word-final positions)

and T4 is performed better at word-final positions," and suggests that "the word-final position is a preferred bearer of a falling contour tone (T4) rather than a rising contour tone (T2)" (Zhang 2013, pp. 119). Using constraint ranking, these propositions can be expressed in Table 14. Again, the asterisks in Table 14, a convention used the OT framework, mean disfavor. So *Fall-I >>*Fall-F means that a falling tone at word-initial positions is worse than a falling tone that occurs at word-final positions.

Applying the same counting method, the percentages of errors in each tone category at either word-initial or word-final positions found in the current study are summarized in Table 15.

The four speakers in the current study made far fewer errors in tone production overall and especially in T1 and T4. So, one can argue that the comparison between the very small percentages in these two tones is not going to yield many meaningful interpretations. But combining the findings in Tables 12 and 13, one can reasonably argue that the rates of errors in T1 indicate that even for a high-level tone, the initial position is more problematic than the word-final position. The falling tone T4 seems neutral with regard to the positional difference. With regard to T2, J, M and F, all made more mistakes ate word-initial positions than word-final positions. X, the least proficient speaker of the four, made more mistakes at word-final positions. Overall, our data presented in Tables 12 and 15 support the TPC claim that the word-initial position is difficult, but it seems to be difficult for all tones rather than for contour tones only, and the rising tones are performed better at word-final positions, not initial positions. If TPC are universal, and their effects exist as detected in previous studies, then it seems such effects are masked by other effects in our speech samples. I will start from the context of error as a possible source of the confounding effects. By combining information of target tone category, substitute tone category, error tone position in a word and the neighboring tone in a word, the error patterns from the speech samples are shown in three Tables 16, 17 and 18.

It is not difficult to notice that more T2 errors show up in the tonal combinations T2-T1 and T2-T4 than in other combinations, and their corresponding substitute forms are HT3-T1 and HT3-T4. Of the four speakers, these two-tone combinations and substitute forms are over 70% of the total T2 errors in the more proficient speakers M (78%), J (70%) and F (70%). X's speech displays more error tone combinations and substitute forms, but errors in these two still compose more than half of the total.

Both HT3 and T1 are found to be substitutes for T2. However, the detailed distribution of the two substitutes shows that HT3 is not only dominant in the number of errors where it is the substitute, but it also appears in more two-tone combinations

Table 15 Error rates with regard to positions in a word in this study

	J		M		F		X		Overall	
	Initial (%)	Final (%)	Initial (%)	Final (%)	Initial (%)	Final (%)	Initial (%)	Final (%)	Initial (%)	Final (%)
T1	3.1	0.6	0	0	3.0	0.8	2.9	0.2	2.1	0.4
T2	13.0	3.6	10.6	1.7	16.6	5.2	18.5	22.9%	14.3	8.2
T4	1.4	0.1	0.8	1.0	1.0	1.0	5.6	7.1%	2.1	2.4

Table 16 Target T2 errors and substitutions in the contexts of neighboring tones in words

Target → Substitute		J	M	F	X	Total
Target tone at word -initial positions						
T2-T4	→ **HT3**-T4	46	25	25	35	131
T2-T1	→ **HT3**-T1	10	25	26	18	79
T2-T2	→ **HT3**-T2	1	1	2	7	11
	→ **T1**-T2	0	0	0	2	2
T2-T3	→ **HT3**-T3	0	1	0	1	2
T2-Tn	→ **HT3**-N	4	1	3	12	20
Target tone at word- final positions						
T4-**T2**	→ T4-**T1**	1	3	1	4	9
	→ T4-**HT3**	3	2	0	18	23
T1-**T2**	→ T1-**HT3**	7	1	6	38	52
T2-**T2**	→ T2-**HT3**	0	0	6	8	14
	→ T2-**T1**	0	0	0	3	3
Others						
T2	→ **HT3**	8	5	4	22	39
	→ **T1**	0	0	0	1	1
Total		**80**	**64**	**73**	**169**	**386**

Table 17 Target T1 errors and substitutions in the contexts of neighboring tones in words

Target → Substitute		J	M	F	X	Total
Target tone at word-initial positions						
T1-T1	→ **HT3**-T1	0	0	1	0	1
	→ **T4**-T1	1	0	0	1	2
T1-T2	→ **T4**-T2	5	0	1	0	6
	→ **HT3**-T2	2	0	0	1	3
T1-FT3	→ **T4**-FT3	1	0	2	0	3
T1-T4	→ **T4**-T4	2	0	0	0	2
	→ **HT3**-T4	1	0	0	8	9
T1-Tn	→ **HT3**-N	0	0	5	1	6
Target tone at word-final positions						
T4-**T1**	→ T4-**T4**	1	0	3	0	4
T2-**T1**	→ T2-**T4**	1	0	0	1	2
Others						
T1	→ **T4**	4	0	0	1	5
Total		**18**	**0**	**12**	**13**	**43**

Table 18 Target T4 errors and substitutions in the contexts of neighboring tones in words

Target → Substitute		J	M	F	X	Total
Target tone at word-initial positions						
T4-T1	→ **T1**-T1	0	0	0	2	2
	→ **HT3**-T1	0	2	3	3	8
T4-T4	→ **T1**-T4	6	6	4	30	46
	→ **HT3**-T4	0	0	0	2	2
T4-T2	→ **T1**-T2	1	0	0	3	4
T4-T3	→ **T1**-T3	3	0	0	0	3
Target tone at word-final positions						
T2-**T4**	→ T2-**T1**	1	2	1	11	15
T1-**T4**	→ T1-**T1**	0	0	0	3	3
HT3-**T4**	→ HT3-**T1**	0	4	3	12	19
T4-**T4**	→ T4-**T1**	0	3	2	28	33
Others						
T4	→ **T1**	2	1	2	14	19
Total		13	18	15	108	154

as the substitute for T2. For speaker J, M and F, HT3 is the only substitute in all two-tone patterns except in T4-T2 target. X's speech is different from the other three speakers' in that T1 is an alternative substitute tone in T2-T2 target pattern and in monosyllabic words. Even so, in X's samples, the number of HT3 substitutes is much greater than the number of T1 substitutes in the T2-T2 targets and the monosyllabic T2 targets. So, it is reasonable to postulate that HT3 is overall the dominant substitute of T2 in the very last stage of tonal acquisition. T1 substitutes may have existed more in earlier stages, but it has gradually disappeared from the speech.

There are not many errors produced for target T1, and the distribution of number of errors in different two-tone combinations is fairly even. When the target T1 is at word-initial positions, it is not clear whether HT3 or T4 is the more dominant substitution form. When the target T1 is at word-final or monosyllabic positions, the substitutes are all T4.

Among all the errors and the substitutions for target T4, one two-tone combination appears to be the most problematic for the four speakers, namely T4-T4. T4-T4 errors account for 53% of all errors in Table 18. Both the word-initial T4 and the word-final T4 have many errors, with the word-initial T4 being more difficult.[5] Speakers J, M and F substitute the word-initial T4 with a T1 in all the erroneous productions, while X has two HT3 substitutions. In word-final T4 errors in the T4-T4 combination,

---

[5]As a reviewer pointed out, in a T4-T4 sequence, native speakers produce the first T4 as HM and not the full falling contour HL, a process called "T4 sandhi." In the current study when the L2 speakers made errors in producing the T4-T4 sequence by substituting the first T4 with T1, they did not articulate an HM but a tone that is judged by transcribers as T1 (HH). It is possible that this error among the L2 learners is related to the perception of the HM in the T4 sandhi.

T1 is the substitution for all four speakers. M and F each have three and two HT3-**T4** combinations in their speech for mispronounced T4-T4 sequences. However, it has to be noted that in these five cases, both the initial and the final tones of the combinations were pronounced incorrectly. They are target T2-T4 combinations and were pronounced as HT3-T1. Their existence is not to be confused with those HT3-T4 combinations for target T4-T4 sequences. Both HT3 and T1 are found to be substitutes for T4. However, T1 is the only substitute when the error occurs at word-final or monosyllabic positions. T1 also appears as the substitute in more cases and in more two-tone combinations when the error occurs at word-initial positions.

Tables 16, 17 and 18 list the tone errors and the substitutions in words and their local tone contexts. When positions and neighboring tones are taken into consideration, there are a few generalizations that emerge from the data. First, tone errors occur in some tonal combinations more often than others. T2-T1 and T2-T4 target combinations are the most difficult combinations for T2, while T4-T4 is the most difficult combination for T4. The only substitute form for T2-T1 is HT3-T1 and the only substitute form for T2-T4 is HT3-T4. In the T4-T4 combination, there are practically two substitution forms: T1-T4 and T4-T1. Second, Tx (x stands for any tone) in HT3-Tx combinations (HT3-T1, HT3-T2 and HT3-T4) has the highest accuracy. There is no error produced in HT3-T1 and HT3-T2 combinations. In HT3-T4 combinations, only speaker X produced errors. The five errors for the HT3-T4 combination in M and F's speech are actually for target T2-T4, in which both the initial and final tones were pronounced incorrectly. Please note the two generalizations above are related in that the (almost) error-free HT3-T1 and HT3-T4 combinations are also the only substitute forms for T2-T1 and T2-T4, which are the most difficult combinations, indicating that HT3 is the least marked tone in this context.

Looking at the frequencies of each error and their percentages in the overall target combinations is not enough. In natural connected spontaneous speech, tones and tone combinations occur in different frequencies. For example, T4 has a higher frequency of occurrence than the other tones, which could mean that two-tone combinations with T4 may naturally occur more often, and more errors are expected even if the possibility of making such errors is the same as making errors for other target tone categories. Therefore, it is necessary to measure the number of errors for each two-tone combination against the total occurrence of the target two-tone combination, and the results of such measurement for T2 errors are given in Table 19 below.

The ratios in Table 19 show how often T2 is pronounced incorrectly in a particular two-tone combination as a percentage of the total production of the target combination. It is very clear that the T2-T1 combination has the highest error rate for all four speakers and is 44.13% overall. T2-T1 target combination does not seem to occur at high frequency in the speech samples, but when they occur, our subjects tend to make more mistakes in producing the T2 in this combination. In fact, **T2-T1** has the highest error rate among all possible two-tone combinations in all four speakers. For J, M and F individually and for all speakers overall, the second highest error rate comes in T2-T4 combination. Although T2-T4 errors are many, there are significantly more correct T2-T4 productions in the speech, resulting in a lower error rate than the one found in T2-T1 combination, but it is still higher than other combinations. Because

Table 19 Target T2 errors and substitutions in the contexts of neighboring tones in words measured as a percentage in the total target combination

Target		J	M	F	X	Total
Target tone at word-initial positions						
T2-T4	Total targets no. of errors	238 46	237 25	165 25	171 35	811 131
	Percentage	19.3%	10.5%	15.2%	20.5%	16.2%
T2-T1	Total targets no. of errors	**37** **10**	**60** **25**	**45** **26**	**37** **18**	**179** **79**
	Percentage	**27.0%**	**41.7%**	**57.8%**	**48.6%**	**44.1%**
T2-T2	Total targets no. of errors	41 1	74 1	44 2	59 9	218 13
	Percentage	2.4%	1.4%	4.5%	15.3%	6.0%
T2-T3	Total targets no. of errors	87 0	80 1	51 0	103 1	321 2
	Percentage	0%	1.3%	0%	1.0%	0.6%
T2-N	Total targets no. of errors	68 4	49 1	32 3	36 12	185 20
	Percentage	5.9%	2.0%	9.4%	33.3%	10.8%
Target tone at word-final positions						
T4-**T2**	Total targets no. of errors	68 4	110 5	62 1	82 22	322 32
	Percentage	5.9%	4.5%	1.6%	26.8%	9.9%
T1-**T2**	Total targets no. of errors	125 7	91 1	97 6	109 38	422 52
	Percentage	5.6%	1.1%	6.2%	34.9%	12.3%
T2-**T2**	Total targets no. of errors	41 0	74 0	44 6	59 11	218 17
	Percentage	0%	0%	13.6%	18.6%	7.8%
T3-**T2**	Total targets no. of errors	70 0	86 0	48 0	60 0	264 0
	Percentage	0%	0%	0%	0%	0%

the numbers of errors in other two-tone combinations are very small, with error rates in the low single digits, it is not very meaningful to compare those numbers in the current study. What is clear is that T2-T1 is the most difficult, followed by T2-T4. Data in Table 19 support the generalization made based on Table 16 that T2-T1 and T2-T4 are the most difficult combinations when it comes to T2 targets. The high error rate in these two two-tone combinations contributes significantly to the overall high rate of error of T2 at word-initial positions and the overall high rate of error of T2 in general. This explains the findings regarding TMS and TPC discussed earlier in the paper.

The examination of the data in the current study can be summarized pertaining to the TMS and TPC. With regard to TMS, T2 has the highest error rate (as a proportion of all T2 target production) and T2 errors make up the largest portion of all tonal errors. T4 has the second highest error rate and T4 errors make up for the second largest portion of all tone errors. T1 comes third but the overall error rate for T1 target is very low. HT3 is the least marked in production both in terms of error rate and its high frequency of occurrence as the substitute tone. With regard to TPC, T1 and T2 are both performed better at word-final positions while T4's performance is about the same in the initial position or the final position. Although this finding is not consistent with the linguistic predictions (Zhang 2004) that T2 is expected to be pronounced more accurately at word-initial positions compared with word-final positions, and the findings in previous studies, a closer look at the tonal context reveals that high error rates in a few particular tonal combinations can explain the higher error rate of T2 at initial positions, indicating that tone-position constraints (i.e., initial versus final positions) interact with tonal combination factors. T2-T1 and T2-T4 sequences together account for more than 34% of the total tonal errors and 54% of the total T2 errors regardless of positional difference. The difference between the findings in the current study and previous studies suggests that when learners progress into higher tone production accuracy, the general effects of TPC are masked by specific effects on certain tone combinations, while the effects of TMS are amplified, as the higher error rates found in T2-T1 and T2-T4 significantly contributed to the higher error rate of T2. The number of high error rate combinations may become smaller as the learners become better with tones, while their proportions in total number of errors increase. This is seen in the performance of the three more accurate speakers J, M and F, whose errors are distributed with higher concentration in T2-T1 and T2-T4. The fourth speaker X has more errors in more combinations. Such differences suggest that these two combinations are the most difficult and may be the last ones to master in the tonal acquisition process. In most previous studies, they were not identified because the subjects in those studies have not reached this stage of IL yet, and the numbers of errors in other patterns, sequences and contexts are large and the overall distribution of errors is (more) widely spread.

3.3 OCP

The third type of constraint that concerns us is OCP. The error rates and patterns in Table 17 indicate that T4-T4 is difficult while those in Tables 16 and 18 indicate that most other identical tone combinations, including T2-T2 and T1-T1, are not difficult. Out of the 403 total target T4-T4 articulations, there are 81 mispronounced ones (20.1%), with either the first T4 or the second T4 pronounced incorrectly (in our data, no T4-T4 combination was pronounced incorrectly in both T4s). This error rate is the highest among all two-tone combinations that involve a T4. Moreover, those T4-T4 errors make up 52.6% of the total 154 errors found in any two-tone combination that involves a T4. In contrast, out of the 218 total target T2-T2 articulations,

there are 30 mispronounced ones (13.8%), with either the first T2 or the second T2 pronounced incorrectly (in our data, no T2-T2 combination was pronounced incorrectly in both T2s). This rate of error is way below the 44.1% rate found for T2-T1 combination. Those 30 T2-T2 errors make up 7.8% the total errors found in any T2 two-tone combinations. And there are only 3 mispronounced T1-T1 sequences out of the 235 targets. Zhang (2013) investigates the occurrences of identical tone combinations in L2 Chinese learners and finds that such combinations are in general fewer in production, and more T1-T1 sequences are found than T4-T4 sequences, which are more than T2-T2 sequences. Zhang argues that the low frequency of such Tx-Tx combinations indicates a higher level of difficulty for these sequences and suggests the effects of OCP. Zhang's method involves a statistical comparison between the actual occurrences of the Tx-Tx sequences and the expected frequencies of occurrences. Due to the nature of spontaneous speech, it is impossible to calculate the expected frequencies of Tx-Tx targets and compare them to the actual occurrences in this study. Speakers may have different preferences of words and word combinations, which result in non-random distribution of tones and tone combinations. But we can compare the numbers of errors in each Tx-Tx to the actual total Tx-Tx targets. Table 20 lists the total number of errors for each Tx-Tx combination and their percentage in the overall Tx-Tx targets. For example, the four speakers made 30 tone errors in T2-T2 combinations, which equals to 13.8% of the 218 total T2-T2 targets. Those 30 errors are 8.7% of the 346 total T2 errors made in any two-tone combination with a T2 in it (including T2-Tx and Tx-T2). The rates of Tx-Tx errors out of Tx-Tx targets show that T4-T4 has the highest error rate (20.1%), followed by T2-T2 (13.8%), and T1-T1 error rate is only 1.3%.

Zhang (2013) conducts another test to investigate possible OCP effects. The error rates of Tx in identical tone sequences (ITC) are compared with the same tone's error rates in non-identical tone sequences (NITC) using Chi-square tests. Such error rates are compared separately at word-initial and word-final positions. For example, T2

Table 20 Tx-Tx errors found in total Tx-Tx targets

	T1	T2	T4
Overall number of Tx	3202	3412	6115
Overall proportion of Tx in the speech	16.9%	18.0%	32.2%
Overall number of Tx-Tx combination targets	235	218	403
Overall number of errors found in the Tx-Tx combination targets	3	30	81
Overall error rate of the Tx-Tx combination	1.3%	13.8%	20.1%
Overall proportion of Tx-Tx errors among all Tx two-tone combination errors	7.9% (3 out of 38)	8.7% (30 out of 346)	60.0% (81 out of 135)

error rate in **T2**-T2 sequences (where the error is at initial positions) is compared with T2 error rates found in T2-T1, T2-T3, or T2-T4 sequences, where T2 is atinitial positions. Then the T2 error rate in T2-**T2** (where the error is at final positions) is compared with T2 error rates found in T1-T2, T3-T2, T4-T2 sequences, where T2 is at final positions. No significant difference was found in the English speakers' data. Table 21 summarizes the test results.

Using the same method to test the data from the four speakers in the current study, we obtained the results shown in Table 22. Statistically significant results are found for T4 at both initial and final positions and T2 only at initial positions. However in

Table 21 OCP test comparing tone error rates found in ITC and NITC contexts from Zhang (2013)

Test items	NITC error rate[a] (%)	ITC error rate	Chi-sq p value	NITC/ITC
T1 initial	24.17	25	0.8805	0.97
T2 Initial	39.17	30	0.1411	1.31
T4 initial	53.75	57.5	0.5595	0.93
T1 final	29.17	32.5	0.5731	0.9
T2 final	77.08	81.25	0.4350	0.95
T4 final	31.25	20	0.0534	1.56

[a]The method of calculation for NITC and ITC error rate in Zhang (2013)
"Tx" refers to the test tones under discussion, "Tx" could be T1, T2, T4
"Ty" means any real stimuli mandarin tone other than Tx. For example, when Tx = T1, then Ty could be T2, T3 and T4
"E" means erroneous tones for target Tx, i.e., any substitute tone for Tx; or the error rates
"N (Tx > Tx/ _Tx)" are the number of times that the learners correctly produced a target Tx as a Tx in the context _Tx. This context is labeled as Tone L = Tx in the test below 102
"N (Tx > E/ _Tx)" is the number of times speakers incorrectly produced a target Tx as an E in the context _Tx (Tone L = Tx)
When Tone L = Tx, the error rates for target Tx in the two contexts of "_Tx" and "_Ty" are
E (Tx/_Tx) = N(Tx > E/_Tx)/ (N(Tx > E/_Tx) +N(Tx > Tx/_Tx)) → ITC context
E (Tx/_Ty) = N(Tx > E/_Ty)/ (N(Tx > E/_Ty) + N(Tx > Tx/_Ty)) → NITC context
When Tone R = Tx, the error rates for target Tx in the two contexts of "Tx_" and "Ty_" are
E (Tx/Tx_) = N(Tx > E/Tx_)/ (N(Tx > E/Tx_) +N(Tx > Tx/Tx_)) → ITC context
E (Tx/Ty_) = N(Tx > E/Ty_)/ (N(Tx > E/Ty_) + N(Tx > Tx/Ty_)) → NITC context
The test compares the error rates of Tx at ITC contexts and NITC contexts respectively

Table 22 OCP test comparing tone error rates found in ITC and NITC contexts in this study

Test items	NITC error rate (%)	ITC error rate (%)	Chi-sq p-value	NITC/ITC
T1 initial	0.7	1.2	0.75	0.61
T2 initial	15.5	6.0	**3.76 (P < 0.01)**	**2.60**
T4 initial	0.8	4.7	**7.26 (P < 0.01)**	0.17
T1 final	0.5	0.0	1.11	∞
T2 final	8.3	7.8	0.26	1.07
T4 final	1.9	3.6	**2.70P < 0.01)**	1.53

the T2 initial position test, the NITC/ITC rate is larger than 1, meaning that there are more errors made in the NITC contexts. This anti-OCP effect is expected because in previous discussion, it is shown that T2-T4 and T2-T1 combination targets have many more errors than T2-T2. The two tests involving T1 both have insignificant results, suggesting that there is not enough evidence for T1-T1 OCP effect. In all, the test results in Table 22 only support the existence of OCP effect in T4-T4 sequence.

4 Discussion and Conclusions

In the previous section, tone error and substitution data were presented and analyzed. Based on the analyses, the first four research questions have been answered.

4.1 Summary of Answers to the Research Questions

The four speakers made more errors in T2, followed by T4, which is followed by T1. T3, including its variants, has the lowest error rate. HT3 is the most dominant substitute form for T1, T2 and T5, while T1 is the most dominant substitute form for T4. The initial position of a two-tone word is more difficult for T1 and T2, but there is not much difference found in T4 error rates with regard to the positions. When tonal errors are analyzed in two-tone combinations, a few sequences stand out as the most difficult. They are T2-T1 and T2-T4 for T2 and T4-T4 for T4. The existence of the high error rate sequences (T2-T1 and T2-T4) could explain the positional differences found for T2 as well as its overall high error rate. The substitute forms for these two-tone combinations also support that HT3 is the least marked tone because the error-free HT3-T1 and HT3-T4 combinations are also the only substitute forms for T2-T1 and T2-T4, which are the most difficult combinations.

The types of errors, the error rates and substitution patterns are very similar among the three more proficient speakers' speech. Overall T2 is the most difficult tone but the proportion of T2 errors in the total number of errors in each speaker's speech negatively correlates to the overall error rate among the four speakers. X, being the least proficient of the four, generally demonstrates the same types of errors, error rates and substitution patterns. However, she made more mistakes in more tone combinations types and the numbers of errors are less concentrated in different sequences.

The findings in the current study pertaining to TMC differ from previous studies where participants were mostly low and intermediate proficiency learners in that T3 does not appear to be difficult in the speech of the four speakers in this study who are advanced/superior level Chinese learners. The current study confirms the previous claim that T2 is among the most difficult tones. Unlike the conclusions made in previous studies, the error study in this research does not support TPC, which predicts that T2 favors word-initial positions and T4 favors word-final positions. A

new discovery of the current study is that among very high proficiency speakers, tone errors occur in higher concentration in only a few combination contexts, based on which I am hypothesizing a common route of IL tonal development: errors are more widely distributed among different tone categories, and in different contexts, but as learners progress, only some combinations remaindifficult for the speakers. This route of IL tonal development cannot be explained by TMC, TPC, OCP, L1 transfer or pedagogical reasons alone. The current study provides strong support to TMC (*FT3 ≫*T2 ≫*T4 ≫*T1 ≫*HT3, Zhang 2013) except that in the current study, FT3 does not appear to be the most difficult. The effects of TPC in our data seem to be weak. OCP effect is only verified in T4-T4 combination but not in T1-T1. T2-T2 combination displays an anti-OCP effect due to the high error rates found for T2-T1 and T2-T4. In the following section of the paper, I will argue that the high error rate found in T2-T1 and T2-T4 sequences, when studied with the corresponding substitution patterns, points to a special configuration of a coarticulation rule as the source of error among these speakers.

4.2   Rule Configuration as the Explanation

One special characteristic of the four subjects in this study is that they are all very advanced in all aspects of pronunciation. It is reasonable to assume that their L2 Mandarin phonological system is very close to that of native speakers'. An examination of the articulation of the respective tone combinations in native Mandarin speaker's speech will provide some clues to the issues under investigation. In the study of Mandarin tones, besides the canonical sandhi rules, there are a couple of so-called tonal coarticulation rules, which manifest themselves in natural connected speech. Wu (1982, 1985), X. N. Shen (1990, 1992) and Shih (1988, 1991) are among the early studies on the tonal coarticulation phenomenon in Mandarin. Chen M. (2000) summarized the findings based on Shih's study and converted the numeric pitch values into tone category representations, shown in Table 23.

The shaded cells in Table 23 are the combination sequences where tonal coarticulation effects are more salient. The superscript + and- represent up or down shift. Shih (1988) captures these effects into four points, which are stated in formal rules in M. Chen (2000), as shown in Table 24.

Table 23 Tonal coarticulation in Beijing Mandarin (M. Chen 2000: 24)

fu-	-ji			
	T1 HH	T2 MH	T3 ML	T4 HL
T1 HH	HH # HH	HH # MH	HH # ML	HH # HL
T2 MH	MH⁻# HH	MH # MH	MH # M⁺L	MH⁻ # HL
T3 ML	ML # HH	ML # M⁻H	=T2 + T3	ML # HL
T4 HL	HM # HH	HM # MH	HM # ML	HM # HL

Table 24 Tonal coarticulation rules in Beijing Mandarin (Chen M. 2000: 24)

1. T4 + anyT	HL → HM/__T (all Ts in Mandarin) starts with either H or M)
2. T2 + T3	ML → M⁺L/MH___
3. T3 + T2	MH → M⁻H /ML___
4. T2 + {T1, T4}	MH → MH⁻/___Hx (x may be H or L)

Table 25 Coarticulation rule 4 in Beijing Mandarin (M. Chen 2000: 25)

Base form	MH. HL (T2 + T4)		MH. HH (T2 + T1)	
Step 1 in derivation	M_. HL	Tone absorption	M_. HH	Tone absorption
Step 2 in derivation	**MH⁻**. HL	**Tone interpolation**	**MH⁻**. HH	**Tone interpolation**

M. Chen (2000) explains that, of the four rules given above, the first three "are all assimilatory in nature," while the fourth rule looks like a dissimilatory rule but can also be interpreted as a tone absorption followed by a tone interpolation process, sketched out in Table 25. According to Chen, the base form for T2 + T4 is MH. HL, which loses the H in MH during the first step derivation when tone absorption rule is applied. Then in the second step derivation, tone interpolation rule applies, and the output is **MH⁻**. HL. A similar process applies to T2-T1 sequence. Treating the fourth rule as a tone absorption and a tone interpolation process allows a unified account for all four coarticulation rules: they all serve to smooth the transition between tonal targets.[6]

Although the function of the four rules is unified in Chen's explanation, the processes involved in the fourth rule are different from those found in the first three rules.

The examination of the L1 Mandarin tonal coarticulation rules allows us to contemplate an explanation for the high error rates found for T2-T1 and T2-T4 combinations in the four speakers' speech in the current study. These two high error rate combinations correspond to the coarticulation forms described by the fourth rule of coarticulation in L1 discussed above. The only substitute form found in incorrectly pronounced T2 in these two combinations is HT3 (ML) while the correct form should be MH⁻, under the coarticulation effect in the speech. Such a correspondence between the errors found in the last stage IL and the coarticulation forms in L1

---

[6]As one reviewer pointed out, besides the phonological account proposed by Chen, the tonal phenomenon also received phonetic explanations (e.g., Xu 2001), which postulate that the downstep effect is due to the peak delay where the target H is realized in the following syllable. Thus, the pitch height of the first syllable was not as high as if it was pronounced in citation. My standing point is that these two theories are both valid on their own premises, and both capture one key issue: the T2-T4 and T2-T1 combinations are realized differently from other tonal coarticulation combinations. When it comes to L2 learners, such a difference causes their L2 phonological system to form different phonological rules/procedures/configurations. A second question from reviewers is whether in L2 learners the errors were indeed "phonological" and not "phonetic," or more precisely articulatory. To answer this question, it will take a few well-designed and well-controlled experiments, where articulatory effects are isolated. That would be my next step.

Table 26 Process for T1-T4 error in the four L2 speakers' speech	MH. HL	Base form
	M_. HL	Tone Absorption
	ML. HL	**Tone Dissimilation**

should not be viewed as a coincidence. I propose the following process to account for the errors and the corresponding substitutes in these two combinations in IL of the four advanced speakers. The difference between the two processes, one in the IL and one in L1, is highlighted in bold in Table 26 below.

The learners hear the lowering of the H in MH⁻ in native speakers' speech, and then constructed and acquired it as dissimilation rather than interpolation. Because all other three coarticulation rules are assimilatory in nature, in which a tone becomes more like the adjacent tone, in the fourth rule, the lowering of H in MH when it is next to a following H misleads the learners to interpret it as a rule of a completely different kind.

This proposal receives support from the substitute forms in the L2 speech. As shown in Table 26, the attested output of the dissimilation is an ML tone[7] and not an MM, though both are possible dissimilation outputs. One phonological difference between ML and MM in the Mandarin tonal system is that the change from MH to ML is tonemic while from MH to MM is not. Shen (1992) proposes three diagnostics to distinguish tonal coarticulation from tone sandhi in Mandarin, among which two are related to our discussion here: first, only assimilation is considered coarticulation but tone sandhi may be both assimilatory and dissimilatory, and second, tone sandhi may effect tonemic change while tonal coarticulation involves only allotonic variations. Shen's criteria have received many challenges from scholars including M. Chen (2000), and we have seen above that the fourth rule of tonal coarticulation is not a straightforward assimilation process. In fact, M. Chen (2000) argues that there is no essential difference between the so-called tone sandhi and tonal coarticulation. However, Shen's two points shed light on the IL tonal system, in which a distinction may exist (for other reasons) and therefore explains why the output of the tone dissimilation rule found in mispronounced T2-T1 and T2-T4 sequences is ML and not MM. I argue that the two distinctions may not hold for tonal systems in general or for the Mandarin native speaker's system, but it reflects a distinction between two types of tonal rules in the advanced speakers IL system (Table 27).

If this distinction exits, then we can anticipate that when they hear the sequence "**MH**⁻-HL," they will process it as a Type 1 rule, because it is dissimilation, and output a tonemic form ML. In fact, ML is the only possible output in the Mandarin system that is both dissimilatory in nature and tonemic. Of course, we do not have to call this distinction as sandhi versus coarticulation, we can call it "Rule Type A" and "Rule Type B." Future experiments will help clarify the nature of the distinction of the two types of processes. In the current study, evidence for this distinction comes from

[7]The two transcribers of tonal data in this study checked many times that the error form that they heard for the target T2 was definitely a low falling, not a MM. The quality of clips does not allow a phonetic study to further confirm the perception, but a future study may continue in this direction.

Table 27 Two types of rules in IL of the four advanced speakers

	Type 1 (Sandhi)	Type 2 (Coarticulation)
Processes allowed	Both assimilation and dissimilation	Only assimilation
Output	Tonemic change	Allotonic variants

the four speakers' acquisition of the other rules. The four subjects have successfully acquired the T3 sandhi rules, the *bu* sandhi rule, the *yi* sandhi rule and some of them also showed competence in using the T2 sandhi rules. All of these sandhi rules are tonemic and include both assimilatory dissimilatory types. They also show little difficulty with the first three coarticulation rules, all which are assimilatory and allotonic. Moreover, they handle the tone-intonation interaction very well and with remarkable fluency. So it is reasonable to suspect the prevalent errors found in T2-T1 and T2-T4 are not due to pure phonetically motivated reasons.

4.3 Constraint Re-Ranking in L2 Acquisition

Lastly, I would like to add some points to the issue of constraint ranking in the OT framework as the explanation of L2 acquisition, the framework employed in Zhang (2010, 2013, 2018). The OT framework is a one-step input–output declarative system, in which markedness constraints and faithfulness constraints interact to select the most optimal candidate as the attested form. Applying the model to SLA, the nature of variability and instability of the IL is captured as the ranking, re-ranking or even the absence of ranking in different conditions (Hancin-Bhatt 2008). Ideally, a complete OT analysis of the L1 system is helpful when studying the rankings of L2 IL, serving as the point of reference for the latter but that does not exist in most cases, including the tonal system of Mandarin. So as Zhang (2013) admits, the rankings/re-rankings analyses only "deal with specific inputs and employ a small amount of the related constraints to illustrate some features of the current inter-language grammars" (Zhang 2013, pp. 186). For example, Zhang (2013) explains the high error rate combinations and their substitute forms using the ranking of a few constraints.

Table 28 only explains why for these L2 learners some (or a lot of) T1-T2 targets were pronounced as T1-T3. We know that the same group also pronounced some (or

Table 28 Tableau for English and Korean speakers' choice of T1-T3 for input T1-T2 (Zhang 2013: 190)

T1-T2	*Rise-F	Id-T
T1-T3		*
T2-T2	*! W	L

a lot of) T1-T2 targets as T1-T2. And the ranking *Rise-F > > Id-T is not necessarily working for other tone combinations such as T4-T2 or T3-T2. It is not necessarily true that the existence of correct T1-T2 or T3-T2 output means that Id-T ranks higher in these other contexts. Other higher ranked constraints may be the reason why T4-T2 or T3-T2 are still chosen as the optimal output even if they violate *Rise-F. A complete OT analysis would have to yield all the correct output and none of the incorrect output for all the single tones and tone combinations. Knowing such limitations, I would cautiously propose the following ranking for the large number of errors found in T4-T4 targets. Among all the identical two-tone combinations (T1-T1, T2-T2, and T4-T4), OCP effect is verified only in T4-T4 combination in the current study, which means a specific form of OCP *HL is ranked higher than the faithful constraint Id-T that requires the input and output tones to be the same. The general OCP that punishes all identical two-tone combination is lower, and so are the subtypes OCP (L), OCP (LH) and OCP (H), because T1-T1 and T2-T2 targets surface as T1-T2 and T2-T2 (in the vast majority cases) and the T3 sandhi is acquired by all four speakers. This ranking is different from the constraint ranking/re-ranking proposed in Zhang (2013), where OCP (L) is promoted to the top of the ranking. The current study confirms that subtypes of OCP constraints (in the L2 learners of Chinese whose native language is English) may go through a re-ranking process separate from the general constraint. The question remains whether OCP (HL) constraint is demoted first and then moves up in the final stage of L2 acquisition process.

The proposed "coarticulation rule configuration" explanation for the high error rates found for T2-T4 and T2-T1 combinations in this paper is formulated in the rule-based framework. Constraints (such as OCP) were proposed in pre-OT phonological theories and were used as explanations for many phonological processes. However in OT, at least in the strict versions of OT, constraints replace all rules. It is possible to reformulate the "coarticulation rule configuration" hypothesis using the OT constraint framework. But this will require the ranking of relevant constraints to account for all the tone sandhi phenomena as well as coarticulation effects shown in Table 22, a research that has not been done in L1 Mandarin phonological studies. Yin (2012) attempts to reach one coherent ranking to account for both T3 sandhi and what he calls T4 sandhi, which equals to the first coarticulation rule in Table 24. Nonetheless, even without such a coherent OT analysis, the OT framework predicts that the difference between the L2 IL discovered in this paper for the very advanced learners and the Mandarin L1 system is possibly that in the L2 IL, there is a higher ranked markedness constraint, which favors the substitute output ML over the L1 output MH$^-$ in T2-T1 and T2-T4 combinations. Coming up with such a comprehensive OT analysis is worth future studies.

To the teachers of Chinese as a foreign language, this study brings them one suggestion. To help the advanced-level learners, contrast practice involving pairs of "T2-T1/T3-T1" and "T2-T4/T3-T4" will help them reduce the errors. The key is to increase the awareness of the T2 end point in the T2-T1 and T2-T4 combinations. Fossilized errors in high frequency words that were acquired earlier require special attention.

In conclusion, this paper investigatesthe errors and substitutions forms of the tone production in four very advanced learners' spontaneous connected speech in Chinese. It is found that the overall ranking of difficultness level of the single tones is T2 ≫*T4 ≫*T1 ≫*T3, therefore supporting the TMC (*FT3 ≫*T2 ≫*T4 ≫*T1 ≫*HT3) proposed in previous studies. T1 and T2 are performed better at word-final positions while T4 is performed similarly at either word-initial or word-final positions. Such TPC effects are different from the theoretical prediction. A close examination of two-tone combinations reveals that the difference is due to the high error rates found in T2-T1 and T2-T4 combinations. OCP effect is found only in T4-T4 combination and not in T1-T1 combination. An anti-OCP effect is found for T2-T2. The different error rates found in different identical tone combination sequences suggest that the subtype OCP (HL) is ranked higher than the generic one and the other subtypes. The high error rates of T2-T1 and T2-T4 are explained as a rule configuration in the IL where the L1 coarticulation rule "MH → MH⁻/ ___Hx" is processed in the learners' phonological system as a tonemic (sandhi) rule. Future study is needed to verify whether the discoveries of this paper are applicable to other very advanced learners of Mandarin. Learners whose native language is not English should be included and longitudinal studies are very much needed to trace the changes. In L1 Chinese phonological study, a complete OT analysis of tone sandhi and coarticulation phenomena will help the L2 researchers pin down exactly the constraint re-rankings that need to take place before the very advanced speakers complete the tonal acquisition.

References

ACTFL, (2012). *ACTFL performance descriptions for language learners.* ACTFL.

Bao, Z. (1999). *The structure of tone.* Oxford University Press.

Chao, Y.-R. (1930). A system of tone-letters. *Le Maitre Phonetique, 45,* 24–27.

Chao, Y. -R. (1968). *A Grammar of Spoken Chinese.* University of California Press.

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

Chen, Q. H. (1997). Toward a sequential approach for tonal error analysis. *Journal of the Chinese Language Teachers Association, 32,* 21–39.

Chen, Q. H. (2000). *Analysis of Mandarin tonal errors in connected speech by English-speaking American adult learners.* [Doctoral Dissertation]. Brigham Young University.

Chaudron, C. (2003). Data collection in SLA research. In C. Doughty and M. Long (Eds.) *The handbook of second language acquisition* (pp. 762–828). Blackwell.

Elliot, C.E. (1991). The relationship between the perception and production of Mandarin tones: An exploratory study. *University of Hawai' I Working Papers in ESL, 10* (2), 177–204.

Goldsmith, J. (1976). An overview of Autosegmental Phonology. *Linguistic Analysis, 2,* 23–68.

Guo, J. -F. (1993). *Hanyu shengdiao yudiao chanyao yu tansuo* [Elucidation and exploration of tone and intonation in Chinese]. Beijing Language Institute.

Guo, L, & Tao, L. (2008). Tone production in Mandarin Chinse by American students: A case study. *Proceedings of the 20*th *north american conference on chinese Linguistics* (pp 123–138). Ohio State University Press.

Hanscin-Bhatt, B. (2008). Second language phonology in optimality theory. In J. Edwards & J. Zampini (Eds.), *Phonology and second language acquisition* (pp. 117–146). John Benjamins.

Hao, Y.-C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics, 40,* 269–279.

Hyman, L. M., & VanBik, K. (2004). Directional rule application and output problems in Hakha Lai Tone. *Language and Linguistics, 5*(4), 821–861.

Kratochivil, P. (1987). The Case of the third tone. In *The Wang Li memorial volumes* (pp. 253–276). Joing Publishing Co.

Leather, J. (1990). Perceptual and productive learning of Chinese lexical tone by Dutch and Englis spehakers. In J. Leather & A. James (Eds.), *New Sounds 90: Proceedings of the Amsterdam Symposium on the Acquisition of Second Language Speech* (pp. 305–341). University of Amsterdam.

McCarthy, J., & Prince, A. (1993). *Prosodic morphology.* Ms., University of Massachusetts, Amherst and Brandeis University.

Miracle, C. (1989). Tone production of American students of Chinese: A preliminary acoustic study. *Journal of the Chinese Teachers Association, 24,* 49–65.

Ohala, J. (1978). Production of tone. In V. A. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 3–39). Academic Press.

Prince. A., & Smolenski, P. (1993). *Optimality theory: constraint interaction in Generative Grammar.* Ms., Rutgers University and University of Colorado.

Shen, X. S. (1989). Toward a register approach in teaching Mandarin tones. *Journal of Chinese Language Teachers Association, 24,* 27–47.

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

Shen, X. N. (1992). On tone sandhi and tonal coarticulation. *Acta Linguistica Hafniensia, 24,* 131–152.

Shih. C. (1988). *Tone and intonation in Mandarin.* Ms., Cornell University and AT&T Laboratories.

Shih, C. (1991). Pitch variation across word boundary. Paper presented at *the Third North America Conference on Chinese Linguistics.* Cornell University.

Sun, S. (1998). *The development of a lexical tone phonology in American adult learners of Standard Mandarin Chinese.* University of Hawai'i Press.

Van de Weijer, V., & Sloos. (2014). The four tones of Mandarin Chinese: Representation and acquisition. *Linguistics in the Netherlands*, 2014, 180–191.

Wang, X. C. (2006). Perception of L2 tones: L1 lexical tone experience may not help. *Proceedings of Speech Prosody*. Dresden, Germany.

Wang, Y., Spence, M., Jongman, A., & Sereno, J. (1999). Training American listeners to perceive mandarin tones. *Journal of the Acoustical Society of America*, *106*, 3649–3658.

White, C. (1981). Tonal pronunciation errors and interference from English intonation. *Journal of the Chinese Language Teachers Association, 16*(2), 27–56.

Wu, Z. J. (1982). Putonghua yuju zhong de shengdiao bianhua. *Zhongguo Yuwen*, 439–450.

Wu, Z. J. (1985). Putonghua sanzizu biandiao guilv. *Zhongguo Yuyan Xuebao*, *2*, 70–92.

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

Xu, Y. (2001). Fundamental frequency peak delay in Mandarin. *Phoetica, 58,* 26–52.

Yang, C. (2011). *The acquisition of Mandarin prosody by American learners of Chinese as a foreign language (CFL.)* [Doctoral Dissertation]. Ohio State University.

Yang, B. (2015). *Perception and production of Mandarin tones by native speakers and L2 learners.* Springer.

Yang, C. (2016). *The acquisition of L2 Mandarin prosody.* John Benjamins.

Yin, H. (2012). A unified account of Mandarin tone 3 and tone 4 sandhi. In *Proceedings of the Third international symposium on tonal aspects of languages.* Nanjing University.

Yip, M. (1980). *The tonal phonology of Chinese.* [Doctoral Dissertation]. MIT.

Yip, M. (2002). *Tone.* Cambridge University Press.

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

Zhang, J. (2004). Contour tone licensing and contour tone representation. *Language and Linguistics, 5*(4), 925–968.

Zhang, H. (2010). Phonological universals and tone acquisition. *Journal of Chinese Language teachers association, 45*(1), 39–65.

Zhang, H. (2013). *The second language acquisition of Mandarin Chinese tones by English, Japanese and Korean speakers*. [Doctoral Dissertation]. University of North Carolina-Chapel Hill.

Zhang, H. (2018). *Second language acquisition of Mandarin Chinese tones—Beyond first language transfer*. Brill.