The effects of type of instruction on the initial stages of L2 perception and production of tones in Mandarin Chinese

Clinton Knight Hendry

A Thesis

in

The Department

of

Education

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Arts (Applied Linguistics) at

Concordia University

Montreal, Quebec, Canada

September 2017

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Clinton Knight Hendry

Entitled: The effects of type of instruction on the L2 acquisition of tones in Mandarin Chinese

and submitted in partial fulfillment of the requirements for the degree of

Master of Arts (Applied Linguistics)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

_____Chair

_____Examiner

Laura Collins

_____Examiner

Kim McDonough

_____Supervisor

Walcir Cardoso

Approved _____by

Chair of Department or Graduate Program Director

_____2017 _____

Dean of Faculty

ABSTRACT

The effects of type of instruction on the L2 acquisition of tones in Mandarin Chinese

Clinton Knight Hendry

This study explored the effects of pronunciation teaching on the L2 acquisition of the four tones that characterize the Mandarin Chinese (MC) tonal system (Qu, 2013):

T1: High and level

T2: High-rising

T3: Low-falling and rising

T4: High-falling

Research indicates that L1 and L2 acquisition of MC tones obeys a developmental sequence suggesting a markedness hierarchy for tones: T1 > T4 > T2 > T3 (Zhang, 2007; where > indicates "acquired before" and "less marked than"). In a study on the acquisition of foreign /s/ + consonant onset clusters (sC; e.g., /st/ in stop) by Brazilian Portuguese speakers, Cardoso (2011a) and Cardoso and Collins (2015) found that students who were taught exclusively the most marked sC acquired all the other clusters, but those who were instructed from least to most marked or all at the same time did not show similar progress. The goal of this study was to apply similar methods to investigate the effects of type of instruction on the acquisition of these tones by L2 learners.

Sixty-two participants, selected for their inexperience with tonal languages, were divided into two experimental groups to be instructed Mandarin's tonal system: one was taught exclusively the most marked tone (T3) during instruction, while the other received instruction in a less marked tone (T4). Both groups were assessed on their ability to perceive and produce the tones they were instructed in and transfer their skills to the other tones. Results indicate that instruction of the most marked tone (T3) is more effective in terms of both accurate perception and production of MC

tones. Our discussion highlights the pedagogical implications of our findings, particularly regarding the teaching of items that follow a developmental sequence in L2 phonology.

Acknowledgements

First and foremost, I would like to thank my advisor, Dr. Walcir Cardoso. It was under his guidance in my first class at Concordia that the beginnings of my thesis emerged, and it has been with his guidance and immeasurable patience that I have made it this far. I would also like to thank my committee members, Dr. Kim McDonough and Dr. Laura Collins, for their valuable time, incredible expertise, thoughtful feedback, and agreeing to help me along during my time at Concordia University. In addition to the above, I would also like to thank Randall Halter for his assistance interpreting the results of my study as I would have been lost without him.

I would also like to thank my colleagues who have consistently been ready to lend a helping hand: June Ruivivar for her amazing last-minute edits, Cynthia Lapierre for her time and patience as I developed my study, and Emily Sheepy for always being there when I need her to be. From outside the department, I would also like to thank Qinghua Song for her teaching expertise and amazing instructional videos.

I wish to also thank my family for helping me reach this point in both life and my academic career, particularly my mother, Wendy Wallace, and father, Neil Hendry who both took the time to read my thesis and give their advice on my grammar and spelling, as all good parents should. I would also especially like to thank my girlfriend, Maggie Gao, for being supportive, thoughtful, and caring during my MA, and hopefully for many years after.

Last but not least, I would like to thank all the administrators and tech support in the Department of Education, Concordia University. It was due to their unfailing guidance and unbelievable patience in the most trying and random of circumstances that I did not fall through the cracks.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

My thesis is a study on the acquisition of Mandarin tones by non-tonal first language speakers; specifically, how varying the learner's exposure to Mandarin tones during instruction affects acquisition. My interest in language acquisition began when I started learning Mandarin while living and working in China in 2010. Learning Mandarin was hard: I remember trying to make myself understood trying to order coffee, telling my taxi driver where I lived, or trying to introduce myself to my new neighbours. At the beginning, I eventually managed to get coffee by ordering in English instead of Mandarin, but communicating with the taxi driver and dealing with my neighbours was less successful. After studying Mandarin for the five years I lived in Beijing, I still struggled with Mandarin's tones and regularly find myself unintelligible to native speakers. Armed with my enthusiasm for learning Mandarin combined with my interest in ESL and language teaching in general, I made the decision to apply for the Masters in Applied Linguistics at Concordia University, Montreal and began my studies in the fall of 2015.

In the first semester of my studies, I found myself applying my own experiences and knowledge from learning Mandarin to my work. The impetus for a thesis came at the end of my first semester of my MA with two end-of-semester projects that centred around the acquisition and markedness of Mandarin tones. My supervisor, Dr. Walcir Cardoso, found the idea of investigating Mandarin tones through phonological theory interesting (e.g., the concept of markedness and its implications), and felt that investigating the acquisition of tones by atonal learners would be an important addition to research in phonology, Mandarin tones, and instructional design.

Further research found that although L2 acquisition of MC has raised some interest among researchers, most of this research has focused on Mandarin's relatively opaque orthography,

sentence processing, word recognition, and speech processing research in psycholinguistics (e.g., Jiang, 2014; Rohr, 2014; Triskova, 2017). Therefore, in response to my interest in the instruction of Mandarin tones and the effects of different types of teaching (i.e., *exposure* to target items, as will be operationalized later), and the fact that tonal L2 acquisition has not received careful attention in research, my thesis aimed to investigate the *initial* stages of acquisition of Mandarin Chinese (MC) tones by speakers of atonal languages such as English and French. I hope to fill this gap in the literature and contribute to the field of L2 tonal acquisition and phonology with my research.

Snapshot of MC tones in the literature

Tones are not only difficult to remember, but also difficult to articulate and acquire for people from any language background, but particularly for speakers of atonal languages (Jiang, 2014; Nguyen & Macken, 2008; Triskova, 2017). However, for all speakers of MC, correct tone production is essential. Tones carry lexical meaning, and correct production is required to differentiate minimal pairs, which might otherwise be homophones. For example, depending on the tone used, the MC word “shui” could mean “sleep” or “water”, the word “ma” could mean “horse” or “mother”, and the word “xue” could be “to study”, “snow”, or “blood”. When tones are confused or not produced at all, the possibilities for communication breakdown increase and the resulting language may become impossible to understand.

Mandarin Chinese has four tones: T1, T2, T3, and T4. As mentioned above, they are changes in pitch that affect lexical meaning, and their correct production is essential for communication in MC (Yip, 2002). For example, although the T1 word *tu* (rabbit) and the T3 word *tu* (earth, ground) share the same segments (and spelling), they are two distinct words with different unrelated meanings. In terms of acquisition, for both L1 and L2 speakers, the body of

research around MC tone acquisition generally agrees that without intervention in a natural setting, MC tones are acquired in a similar order: T1 > T4 > T2 > T3 (Zhang, 2007, Li & Thompson, 1976; where > indicates "acquired before"). More details will be provided on MC tones in Chapter 2.

Considering the above, from a research perspective, tones constitute interesting phonological features to investigate because they are assumed to follow a developmental sequence or order of acquisition (T1 > T4 > T2 > T3) within which certain tones (the easy, less marked tones) are acquired before others (the harder, more marked tones; e.g., Nguyen & Macken, 2008; Qu, 2013; Zhang, 2007; Li & Thompson, 1976). This pattern corresponds to Eckman's (1977) Markedness Hypothesis, which proposes that some linguistic structures (the harder ones) can only be learned after others (their easier counterparts). Consequently, in this study, I assume that there is a predictable universal tone acquisition order (based on studies by Nguyen & Macken, 2008; Qu, 2013; and Zhang, 2007), which is consistent across all tonal languages, as will be discussed later.

Orders of acquisition or markedness hierarchies have been researched across many languages (see Shirai, 1997, for some examples) and different theories have been put forward regarding the best way to take advantage of these hierarchies or developmental sequences for pedagogical purposes. For instance, Pienemann (1989, 2005) argued that students should only be taught one level higher than learners' current level. In other words, if language learners know *n*, then they should be taught *n*+1 next. Another proposal, the Projection of Markedness Hypothesis by Zobl (1983, 1985), argued that teaching the most marked structure, often the last naturally acquired one due to its difficulty, should be the only structure deserving of an instructional focus for the sake of efficiency. Other less marked structures would eventually be acquired because they are easier, or learners would be able to generalize their knowledge of the most marked structure to

others less marked ones, without instruction. Both theories have been used in previous studies to inform instructional design (e.g., Lightbown & Spada, 1990; White, Munoz & Collins, 2007), but the majority of the available research on teaching with markedness hierarchies has focused only on morphosyntactic structures. Using different instructional methods that vary exposure in learners' acquisition of phonology still has room for exploration.

Cardoso (2011a), in one of the only studies that examined the effects of instruction on L2 phonological acquisition, compared three instructional methods on the acquisition of sC onset clusters (/s/+ consonant sequences as in /st/op, /sl/eep) by Brazilian English learners. He found that the participants using a method inspired by Zobl's (1983, 1985), focusing exclusively on the most marked sC cluster (/st/), were best able to learn all sC clusters (/st, sn, sl/). The groups that received piecemeal instruction (from least to most marked) or were taught all sC forms at once underperformed in oral production (see also Cardoso & Collins, 2015, for similar results). Based on these results, it is possible that Zobl's proposal might be an effective method for teaching developmental sequences in the realm of phonology.

However, more research is needed to verify whether the results found in Cardoso (2011a) and Cardoso and Collins (2015), supporting Zobl's Projection of Markedness hypothesis, can be generalized to other phonological phenomena and across other languages. In addition, we need to address one of the limitations of their study: were their results a consequence of an instructional focus on a single item (the most marked) throughout the experiment? Would the results be similar if the students had been taught exclusively the easiest, less marked form because instruction of a single form is more effective than instruction of multiple forms no matter their markedness? Following the above rationale, one of the goals of my thesis is to address this limitation by focusing on the two ends (easy vs. hard) of the developmental sequence that characterizes MC tonal

development. Recall from before, the predicted order of acquisition for all MC tonal learners is: T1 > T4 > T2 > T3 (Zhang, 2007, Li & Thompson, 1976; where > indicates "acquired before", "less marked than", and likely "easier"). Therefore, as the rationale for this study is based on Zobl's hypothesis that instruction focus on the most marked form might lead to the learning of other less marked forms, but also to target only single phonological structures to address the limitations in Cardoso (2011a) and Cardoso and Collins (2015), ideal targets for study are one of the two easier tones (T1, T4) and one of the two more difficult tones (T2, T3). The reasoning behind which tones are selected for comparison (i.e., T4 and T3) are discussed in Chapter 2.

In summary, with the growth in popularity for learning MC as a foreign language, now is an opportune time to press for more research into L2 MC, and because tones are some of the most difficult phonological features of MC to acquire (Triskova, 2017), they constitute an interesting target for investigation. Furthermore, the consistent order of tone acquisition across both native and non-native speakers in MC (e.g. Hao, 2012; Nguyen & Macken, 2008) makes tonal acquisition an ideal candidate for testing theories such as Zobl's Projection of Markedness (1983, 1985) for the teaching of developmental sequences. My study contributes not only to the field of tonal acquisition, but also to the debate over the instructional methods explored in Cardoso's and Cardoso and Collins' studies involving a new target language and a new phonological feature, and to the greater body of research surrounding the teaching of marked structures that constitute development sequences.

Because this is manuscript-based MA thesis, Chapter 2 consists of a research paper ("a full submittable draft of a manuscript", as indicated in the MA thesis guidelines) in which parts of this chapter are repeated in condensed form. Accordingly, the following chapter is a manuscript of the study proposed for my thesis.

Chapter 2

Introduction

Tones are changes in pitch that affect lexical meaning. For learners from atonal backgrounds, they are considered particularly difficult to perceive and articulate (Jiang, 2014; Nguyen & Macken, 2008; Triskova, 2017). In Mandarin Chinese (MC), correct tone production is essential as tones carry lexical meaning and are required to differentiate minimal pairs. When tones are confused or not produced at all, spoken language quickly breaks down and becomes unintelligible. However, research into the L2 acquisition of MC has focused mostly on Mandarin's relatively opaque orthography, sentence processing, word recognition, and speech processing research in the field of psycholinguistics. L2 tonal acquisition has not received as careful attention from researchers (Jiang, 2014; Rohr, 2014; Triskova, 2017). This study addresses the L2 acquisition of MC tones by exploring the effects of different instruction types.

Mandarin Chinese itself has four tones: T1, T2, T3, and T4. As mentioned above, they are changes in pitch that affect lexical meaning, and their correct production is essential for communication in MC (Yip, 2002). For example, although the T1 word *tu* (rabbit) and the T3 word *tu* (earth, ground) share the same segmental structure, they are two distinct words with different unrelated meanings. In terms of acquisition, for both L1 and L2 speakers, the body of research around MC tone development generally agrees that, without intervention, in a natural setting, MC tones are acquired in a similar order: T1 > T4 > T2 > T3 (Zhang, 2007, Li & Thompson, 1976; where > indicates "acquired before"). This order of acquisition, or developmental hierarchy, constitutes an interesting target for investigation into how tones are acquired by L2 MC learners, as will be discussed next.

Orders of acquisition or markedness hierarchies have been researched across many languages (see Shirai, 1997, for some examples) and different theories have been put forward regarding the best way to take advantage of these hierarchies or developmental sequences for pedagogical purposes. However, the majority of available research on teaching involving markedness hierarchies has focused on morphosyntactic structures (e.g., Lightbown & Spada, 1990; White, Muñoz & Collins, 2007). Cardoso (2011a), in one of the only studies that examined the effects of instruction on L2 phonological acquisition, compared three instructional methods on the acquisition of sC onset clusters (/s/+ consonant sequences as in /st/op, /sl/eep) by Brazilian English learners. He found that the participants using a method inspired by Zobl's (1983, 1985), focusing exclusively on the most marked sC cluster (/st/), were best able to learn all sC clusters (/st, sn, sl/). The groups that received piecemeal instruction (from least to most marked) or were taught all sC forms at once under-performed in oral production (see also Cardoso & Collins, 2015, for similar results).

Inspired by Cardoso (2011a) and Cardoso and Collins (2015), this study assessed two groups of participants: one that was instructed only in the most marked tone, T3 (most difficult to acquire; last acquired in a natural setting), and a less marked tone, T4 (less difficult to acquire; acquired earlier in a natural setting). In accordance with their findings, who found that teaching of the most marked (most difficult to acquire) structure leads to more accurate production of structures within said hierarchy, participants in our study were tested on their ability to accurately perceive and produce all four MC tones (T1, T2, T3, and T4) to test their ability to generalize their knowledge to new, unfamiliar structures.

To contextualize the current study and define its scope, goals and predictions, the following section defines tones, describes them in terms of markedness, and discusses instructional methods that make reference to this concept.

Literature Review

Tones

Tones are essential to learning tonal languages such as Mandarin Chinese (MC). The changes in pitch that characterize tone assignment affect lexical meaning (similar to /p/ and /b/ in English; pin vs. bin respectively), thus indicating that their correct pronunciation is required for mutual intelligibility. Tones are present in several languages, including MC, Swahili, Vietnamese, and Yoruba. Yip (2002) estimates that 70% of the world's population speaks a tonal language, and MC, with over a billion speakers, likely accounts for the largest population of tonal speakers.

In most tonal languages, regardless of the number of tones, tones are categorized in terms of level and contour. Figure 1 illustrates *level* tones in a 5-point pitch scale for MC tones established by Chao (1968). Observe that a level tone starts and ends at the same pitch level, and for that reason they are often represented orthographically as 55, 33, 11, etc. *Contour* tones, on the other hand, have a change in pitch within a single morpheme (most often a vowel) and, accordingly, they are illustrated in writing as 35, 214, 51, etc. As mentioned earlier, tones are essential for mutual intelligibility because they affect lexical meaning. For example, in MC, the *level* tone 55 in "chi" means 'to eat', while the same word in the *contour* tone 35 means 'to grasp' (See Figures 1 and 2 for an illustration of these two tones).

Tonal languages can have a maximum of five contrastive level tones, although many have fewer (Maddieson, 1978). A similar observation is made by Li and Thompson (1976), who use a

5-point pitch scale for describing tones. This is the scale adopted to describe MC tones, as will be discussed next.

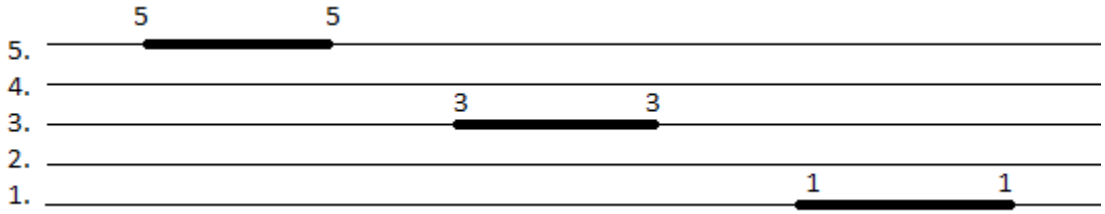

Figure 1. Level tones in a 5-point pitch scale (adapted from Chao, 1968 and Maddieson, 1978).

Mandarin tones

Using the system described in the previous section to illustrate tones and their level/contour forms, MC can be described as a four-tone language that conforms to the pictorial representations illustrated in Figure 2.

T1: high tone (55)
T2: rising tone (35)
T3: dipping tone (214)
T4: falling tone (51)

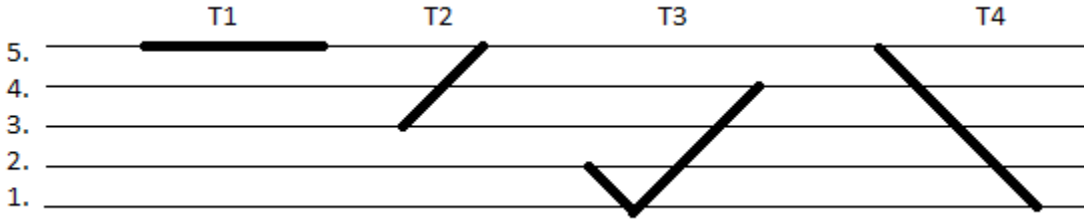

Figure 2. Mandarin Chinese tones on a 5-point pitch scale (adapted from Chao, 1968 and Maddieson, 1978)

In addition to the four traditional or standard tones, there is a neutral tone (T0) which is carried by reduced syllables, in particular certain suffixes, particles, and the second syllable of many words (Li & Thompson, 1976). There is debate as to whether T0 should be recognized as a tone by itself, or whether T0 should be a reduced version of any of the four recognized tones (Qu, 2013). Qu argues that the neutral tone should be considered a full tone in its own right, but as it is the least

marked tone (Maddieson, 1978) and consequently requires the least instruction for acquisition as it is the first acquired in a natural setting, I will refer to it as the "neutral tone" to differentiate it from the four traditional tones (T1, T2, T3, T4).

As mentioned above, to produce any morpheme either as a word or as part of a word in a way that a Mandarin speaker could comprehend, the speaker must also include one of the four lexical tones illustrated in Figure 2. For example, the morpheme *ma* can be used with any of the Mandarin tones. One can quickly see why accidentally calling somebody's mother a 'horse' could be a problem whether in diplomatic circles or with new acquaintances, or why calling a horse 'mom' might lead to some confusion on the farm.

Ma[T1]: 'mother'
Ma[T2]: 'hemp'
Ma[T3]: 'horse'
Ma[T4]: verb; to verbally abuse, swear, call names

Due to their phono-semantic complexity and their difficulty to be perceived and produced (Jiang, 2014; Nguyen & Macken, 2008), tones are difficult to acquire for learners from atonal first languages (L1s) and, to a lesser degree, even for native speakers of MC (e.g., Li & Thompson, 1976). The latter will be the focus of next section in which we establish a developmental hierarchy for tones in Mandarin.

L1 Acquisition of Mandarin Tones

The body of research surrounding MC tonal acquisition generally agrees on the following order of acquisition: T1 > T4 > T2 > T3 (Zhang, 2007; where > indicates "acquired before" and consequently "less marked than"). However, there can be some individual variation. Generally speaking, native MC-speaking children will produce tones correctly before they are able to accurately produce segments such as vowels and consonants and syllables (Li & Thompson, 1976; Zhu & Dodd, 2000), but they do not acquire all tones at the same time. Li and Thompson found

that children have difficulty with T2 and T3, especially when compared with T1 and T4. In their study of MC native speaker children aged 1 to 3, they showed pictures to children and asked them to say the word the picture represented. Based on their productions, the authors divided tonal development in children into four stages. Starting at Stage 1, the child is only able to produce a few words, and they will predominantly use the high tone, T1, or the falling tone, T4, whether or not that tone is the correct tone for that morpheme. At Stage 2, however, the child can produce all four tones with some accuracy, but there might be confusion between T2 (rising tone) and T3 (dipping tone). At Stage 3, the child produces longer strings of 2-3 words still with some T2/T3 confusion, and Stage 4 has the child producing longer sentences with almost no perceptible errors. In the end, despite having four developmental stages from beginning to full acquisition, children as young as 3 years old were able to produce accurate tones for all lexical productions. This accuracy in tonal production is not specific to MC; children have also been found to correctly produce tones at around the same age in other tonal languages, such as Lao (Westermeyer & Westermeyer, 1977). The findings regarding tonal acquisition in MC illustrate what is considered the tonal markedness hierarchy (i.e., T1 > T4 > T2 > T3), which is often described using markedness as an explanatory construct, as will be addressed in the following section.

Tonal Markedness and its developmental sequence

Markedness theory assumes that certain linguistic items can be ranked in a "markedness hierarchy" (Shirai, 1997), reflecting their order of development, as was shown for MC tones in L1 acquisition. The hypothesis behind the concept of markedness (e.g., Eckman, 1977; Zobl, 1983) predicts that the order of acquisition of developmental sequence items will follow a pattern in which easier-to-acquire unmarked linguistic structures are learned before their more marked counterparts (Shirai, 1997). This order has also been observed in tones. In MC, as discussed above,

tones are acquired in the following order and, accordingly, they constitute a markedness hierarchy: T1 > T4 > T2 > T3 (Zhang, 2007; where > indicates "acquired before" and consequently "less marked than").

Maddieson (1978) describes three factors for indicating markedness in tones: frequency (i.e., a less frequent tone is more marked), dominance and assimilatory processes over other tones (i.e., tones left unchanged are more marked), and the fact that neutral tones are less marked than all other tones. Using the above rules, he argued that contour tones are more marked than level tones, rising tones are more marked than falling tones, and complex contours (involving two or more directions in pitch, e.g., T3) are more marked than simple contours (involving only one change in pitch, e.g., T2, T4). This translates in MC as the following generalizations: T3 (complex contour) is more marked than T2 (rising simple contour), T2 is more marked than T4 (falling simple contour), and T4 is more marked than T1, the level tone. This is consistent with the hierarchy proposed by Zhang (2007), discussed previously, and the literature on the markedness of tones and difficulty of acquisition (e.g. Yip, 2002; Zhu & Dodd, 2000; Hao 2012). As will be discussed below, this hierarchy is also supported by what is observed in L2 development: T1 > T4 > T2 > T3.

L2 Acquisition of Tones

Concerning the L2 acquisition of MC tones, there has been only limited research and, surprisingly, that research has predominantly concentrated on the perception of tones (Rohr, 2014; Wang et al., 1999). What research on L2 acquisition of MC tones there is, particularly invloving *atonal* language speakers, shows a very similar order to that of L1 learners (Wang et al., 1999; Nguyen & Macken, 2008; Hao, 2012). However, there is some debate on which tones are acquired more easily, and in what order. For example, Hao (2012) examined differences in Mandarin tone

acquisition between English and Cantonese learners of Mandarin in three tasks: mimicry, identification, and reading. She found that the learners were more accurate in identifying T1 and T4 than T2 or T3, but they were similarly accurate within those sets. That suggests that it is possible for T4 to be perceived before T1, and that T3 may be easier to perceive than T2 because, within those sets, the author did not find a significant difference between which tone was perceived earlier and more accurately. Hao also found that learners may reach 90% accuracy in perceiving that less marked tones are different from more marked tones, but they had some difficulty in perceiving the difference between T1 and T4, and quite a lot of difficulty perceiving the difference between T2 and T3. However, one should note that the group sizes in Hao's study were small (10 English speakers and 10 Cantonese speakers), which may explain some of the discrepancies and individual differences observed. Precisely, the small sample size might have inflated any differences and may not illustrate a group pattern. Additionally, the study included two different tonal languages (Mandarin and Cantonese), thus raising the possibility that the L1 may have exerted an influence on the reported outcomes on the Cantonese speaking participants, as has been observed in Saito and Wu (2014).

Nguyen and Macken (2008) conducted a similar study with English L1 learners of Vietnamese tones. They found that more marked tones, as defined and described here, were more difficult to produce than less marked tones. That is, contour tones were more difficult than level tones, and rising tones were more difficult than falling tones. This would imply that T4 (51, a contour tone) would be more difficult than T1 (55, a level tone), and T2 (35, a rising tone contour tone) would be more difficult than T4, a falling contour tone. They also found the falling-rising tone particularly difficult to acquire. However, unlike Mandarin, Vietnamese contains two falling-rising tones, one of which, a glottalized falling-rising tone, was produced with high accuracy (thus

contradicting the rules of markedness). Nguyen and Macken speculate that the glottalization aided the learners, and they also point out that their small sample size ($N = 6$) might have affected the results.

The reasons for this observed difficulty in tonal acquisition are not clearly understood. Nguyen and Macken list possible contributing factors that may affect English L1 learners of Mandarin, including: the use of lexical pitch in Mandarin and non-lexical pitch in English; a wider range of pitch in Mandarin than English; interference from English intonation; and a difference in emphasis use (in Mandarin, emphasis is produced with volume while in English it is produced with higher pitch and longer syllable duration). Nguyen and Macken tested nine hypotheses on factors affecting tone production in English learners of Vietnamese and found that tone type had the largest effect. Recall that Hao (2012) tested English L1 learners of Mandarin in tonal identification, mimicry, and reading, and were found to be significantly more accurate in mimicry than reading and identification. However, participants were still more accurate in perceiving/identifying T1 and T4 than T2 and T3, implying that tone type still had a large effect on acquisition, which confirms Nguyen and Macken's (2008) findings.

The research outlined above offers considerable support for the tone markedness hierarchy established earlier. For this reason, in this study, it is assumed that the traditional hierarchy of tonal markedness (T1 > T4 > T2 > T3; Zhang, 2007) is accurate and therefore holds for L2 learners. Accordingly, and based on the existing literature, some individual variation between participants, specifically concerning T1/T4, and T2/T3, is expected. This consistency in the markedness hierarchy allows us to examine the effects of exposure to linguistic items that are acquired in stages, such as instruction based on Zobl's Projection of Markedness hypothesis (1983, 1985).

Zobl's Projection of Markedness Hypothesis

A theory that takes advantage of the acquisition of linguistic structures through stages is the Projection of Markedness hypothesis, conceptualized by Zobl (1983, 1985) in an attempt to explain the average learner's acquisition of a language in a relatively short period of time and with comparatively little L2 input.

In his 1983 paper, Zobl argued that when a learner receives input regarding a certain linguistic structure (W, X, Y), they will also acquire information regarding the structure that was not available in the input (Z). In other words, when receiving input W, X, and Y, the learner is able to project to Z. Zobl also points out that this is a method to identify unmarked structures as he believes that Z must be unmarked, for it can be acquired (or often is acquired) without any direct input in the primary data. Two years later, Zobl (1985) applied his theories in studies involving English learners' use of the possessive determiners *his* and *her*. First, he proposed that using determiners with humans, [+human] (e.g., his mother/her father) was more marked than using determiners with inanimate objects, [-human] (e.g., her pencil, his phone). He also identified "her" as being more marked than "him" since "him" is often overgeneralized by English L2 learners. He then had one participant group instructed only in using determiners with humans, the [+human] group, and one that was only instructed in using determiners with non-humans, the [-human] group. His analyses indicated that the [+human] group was able to project their more marked knowledge to the unmarked [-human] determiner use. In contrast, the [-human] group, although able to project a small amount to the [+human] group, were overall not nearly as capable. In summary, Zobl's Projection of Markedness Hypothesis suggests that when learners are taught (and acquire) more marked structures, they are able to project that knowledge to less marked structures, but not vice versa.

Note that Zobl's hypothesis does not contradict Pienemann's Processability theory entirely, but it does promote a different target for instruction. Zobl proposes targeting the most marked stage of a linguistic structure while Pienemann proposes targeting only the stage that is slightly more advanced (or more marked) than the learner's current interlanguage stage. In this study, the focus will be on the two ends that characterize the tonal hierarchy adopted in the study. Will a focus on the most marked T3 tone lead to the acquisition of its less marked forms, or will a focus on a less marked structure (T4) lead to the acquisition of the other forms?

It should be mentioned that Zobl (1983, 1985) specifically looked at grammatical structures. However, his theory has recently been tested in phonological acquisition (Cardoso, 2011a; Cardoso & Collins, 2015), as will be discussed below.

The Effects of Instruction on the Acquisition of an L2 Phonological Structure

Cardoso (2011a) and Cardoso and Collins (2015) looked specifically at different types of instruction (or orders of exposure to target forms) in the development of phonological syllable structure. Cardoso (2011a) targeted the acquisition of English-like homorganic sC clusters (i.e., sharing the same place of articulation) among Brazilian Portuguese speakers, based on the hypothesis that sC clusters are involved in a markedness hierarchy: /sl/ is the least marked (easier to acquire) form and /st/ is the most marked (difficult to acquire): /sl/ > /sn/ > /st/. Three groups were instructed in an artificial language to test three different instructional methods or type of exposure: one based on Pienemann's Teachability Hypothesis that argues for instruction based on the developmental sequence seen in a natural setting, another based on Zobl's Projection of Markedness hypothesis (as discussed above), and a third based on Ammar and Lightbown's (2004) reluctance to recommend instruction via piecemeal exposure. Cardoso's and Cardoso and Collins' results suggest that those exclusively taught the most marked sC cluster, /st/, were able to produce

all three clusters with similar or higher accuracy than the other two groups, thereby supporting Zobl's Projection of Markedness hypothesis. However, one limitation in these studies is that there might be another explanation for why the group that was taught /st/ exclusively did better than the other groups: the participants might have improved significantly and outperformed those in other experimental groups simply because they received instruction on a single structure, throughout the duration of the study (i.e., /st/). It is possible that the lack of a prolonged and concentrated focus on a single form could have hindered overall acquisition, as the participants were not cognitively prepared for such complex instruction (e.g., Pienemann, 1989).

The current study addresses this limitation by exploring two types of exposure to (teaching of) Mandarin tones. Specifically, to address the questions remaining from Cardoso (2011a) and Cardoso and Collins (2015) and to explore Zobl's Projection of Markedness hypothesis, this study tested the effects of two types of exposure to MC tones on its acquisition in both production and perception by atonal L1 speakers. More specifically, we investigated the effects of two types of explicit teaching on the development of MC tones: one that focused exclusively on the most marked tone (T3), and another that targeted a less marked tone (T4).

Research Questions

The purpose of this study was to examine the effects of two types of exposure to phonological forms, as defined earlier, on the acquisition of MC tones by atonal L2 learners. Specifically, it addressed the following research question:

- Which type of exposure is most effective for the *perception* of MC tones and by speakers of atonal L1s, and
- Which type of exposure is most effective for the *production* of MC tones by speakers of atonal L1s:

- one that focuses on the most marked T3 form, or
- one that targets the less marked T4?

For target of instruction, we selected T3 (the most marked tone) and T4 (the second least marked tone). The least marked tone (T1) was not used because it was felt that, as it does not require a change in pitch, it was not as suitable a target for comparison. In addition, as reported earlier, there was the possibility that participants might confuse T1/T4 and T2/T3. Therefore, T4 seemed an ideal choice as it requires a change and pitch, and participants were not likely to confuse it with T3. In other words, we believe that these two tones were sufficiently different in terms of markedness for the purposes of our study.

Based on Cardoso's (2011a) and Cardoso and Collins' (2015) findings, it is hypothesized that the group utilizing instruction informed by Zobl's Projection of Markedness Hypothesis (i.e., with an exclusive focus on T3) will perceive and produce T1 (the level, non-contour tone), T2 (the rising tone), T3 (the falling and rising tone), and T4 (the falling tone) more accurately than the participants learning the fourth Mandarin tone (T4) exclusively.

Method

This study employed a quasi-experimental between/within-groups pretest/posttest design with two experimental groups: experimental group 1 (Group G3) learned the most marked Mandarin tone (T3), while experimental group 2 (Group G4) learned the second least marked Mandarin tone (T4).

Participants

The participants were 62 atonal language speakers ranging in age from 18 to 62 years from a variety of L1s backgrounds (a copy of the questionnaire can be found in Appendix A; a summary of the demographic information can be found in Appendix B). All participants had at least some

knowledge of a second language with most speaking three or more. This linguistic profile is typical in Montreal where the study was conducted. However, to control for previous knowledge of tones, we ensured that no participant had learned or spoke a tonal language. This was confirmed with each participant through email (the call for participants also indicated this requirement), verbally before they began the study, and once more in the demographics questionnaire.

Participants were remunerated 20 dollars per session for their time.

Materials

Teaching sessions. The teaching methods utilized by the instructor followed those used in Cardoso's (2011a) and Cardoso and Collins' (2015) study in a classroom setting, with the exception that the instruction in this study was video-based. The videos were prepared by an MC language instructor and covers all the teaching material (one video for each group). The MC instructor was a native MC speaker with 7 years' teaching experience and with a certificate in teaching Mandarin as a foreign language. One video was created for each treatment group (see a screen shot of the videos in Figure 3 and the scripts for both videos in Appendix C).

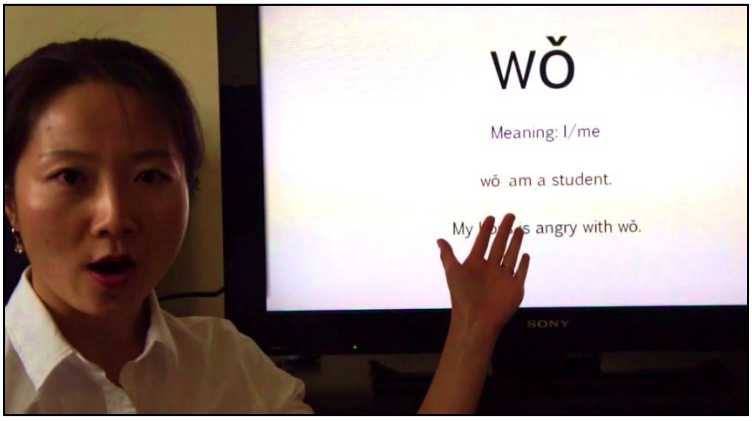

Figure 3. Screenshot of instruction video for G3; [wo3] "I/me"

The videos and teaching sessions were exclusively in English, except for the target MC words and associated pronunciations, and focused on teaching vocabulary containing the target tones: T3 and

T4. Accordingly, instruction obeyed the following procedures (based on Thornbury's 2002 recommendations for teaching vocabulary and associated word knowledge, including pronunciation):

1) Introduction of the Mandarin morpheme's pinyin in the instructional video (pinyin refers to a phoneme-based Romanised spelling system that uses sound-to-letter rules similar to those of English; e.g., *ba* was used instead of Chinese characters);
2) Pronunciation instruction followed by repetition from the participants, based on a model provided by the instructor;
3) Dictations utilizing the vocabulary target item; and
4) Word retrieval activities (e.g., "How do you say ____ in Mandarin?") and using the Mandarin word in an English sentence (e.g., wo3 → "I"; [wo3] am a student).

The participants were allowed to ask the researcher questions at any time.

Each group's instructional video consisted of 10 vocabulary items and lasted approximately 20-30 minutes. The G3 video instructed the participants in ten commonly used T3 vocabulary items, while the G4 video instructed the participants in 10 commonly used T4 vocabulary items. In addition, to reduce any differences in instruction that might impact acquisition, the words presented to each group (G3 and G4) were minimal pairs taught in the same order. For example, the first word three words taught to the G3 were [wo3], [ba3], and [da3], while the first three words taught to the G4 were [wo4], [ba4], and [da4].

Instruments

Questionnaire. The questionnaire can be seen in Appendix A. It asked for basic demographic information (age, gender, education, English and French fluency), and then to list any languages that the participants could read, write, speak, or understand, and/or in which they

had received instruction to confirm whether or not they met the language requirements to participate in the study. In addition, the questionnaire asked if the participant had any particular musical ability or how able they were to mimic accents. These last two questions were added to see if either musical ability or strong imitation skills might correspond with the ability to either perceive or mimic tones (see Delogu, Lamis, and Belardinelli, 2006 for the rationale behind the inclusion of this question). After finishing the questionnaire, each participant was briefed on the exercises (discussed in detail later) before they initiated the testing session.

Multiple-choice question test. The first perception test was a multiple-choice question test (MCQ) in which the participant was asked to listen to 24 MC monosyllabic words played through a PowerPoint presentation, one at a time. Each tone was equally represented and was heard 6 times (24 total). After hearing each word with its associated tone, the participant was asked to choose the tone they believe they heard from the answer sheet provided.

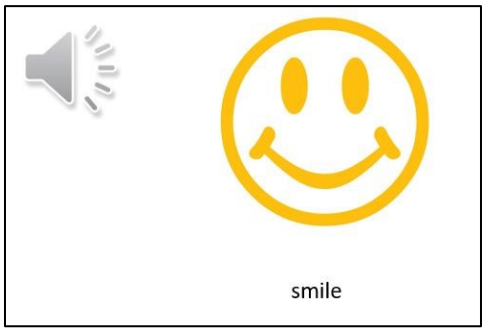

Figure 4. Example of PowerPoint slide used for testing with accompanying MC question

ABX test. The second perception test was an ABX test with 24 items. This test required that the participant listen to three monosyllabic words: A, B, and then X. A and B were minimal pairs in that their only difference was the tone used, and X was a copy of either A or B. The

participant was asked to put a check mark on their answer sheet beside either A or B depending on which they believe was a reproduction of X.

Production test. The third and final test was an oral production test. Similar to the MCQ test, the participant listened to 24 monosyllabic words (one at a time) from a native MC speaker. Following each MC native production, the participant was asked to repeat the entire word (i.e., the segments and its prosodic content, its tone) to the best of their ability. Note that some target vocabulary items used in the tests were segmentally and syllabically similar to the ones used in the teaching sessions; however, they did not contain the same tone assignments. That is, items used in instruction were not used in testing, although there were some minimal pairs.

As mentioned earlier, vocabulary items used in testing were not used in instruction and vice versa. However, the procedures were essentially the same for both the pre- and posttests, and the same MC vocabulary items were used in both, but not in the same order to reduce the possibility that the participants' memory might be improved by seeing the same items in the same order as on the pretest.

Procedure

This study was a "one-shot experiment" to examine the participants' initial stages of MC tone acquisition, and consequently, the pretest, instruction, and posttest took place within the same session. The design of the study can be seen in Figure 5.

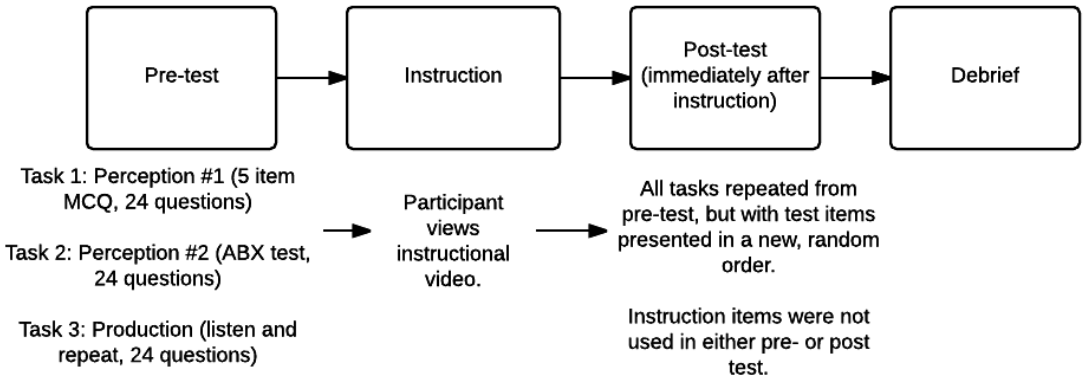

Figure 5. Study design.

Each participant took approximately 1 hour to finish all three sections. All 62 participants were randomly assigned to either G3 (exposure to only T3 during instruction) or G4 (exposure to only T4). The independent variable (IV) was the type of teaching (focus on more or less marked tone), and the dependent variable (DV) was the participants' perception and production of tones, as measured in the three tests described above.

Teaching Sessions. The treatment was conducted using instructional videos created by a native MC language instructor, which focused on both perception (listening) and oral production of MC vocabulary. The sessions were individualized with one researcher and one participant present at a time to better accommodate their schedules and to limit the exposure the participants might receive from hearing others practice the MC lessons. At each session, the researcher, an upper-intermediate level Mandarin speaker, was present to play the video, provide and guide the participants through the assessments, and answer any questions.

Testing. The study followed a pre-/posttest design, as illustrated in Figure 5. Before the pretest, the participants were asked to read and sign a consent form that outlined the study and how their data will be used followed by filling out the demographics questionnaire seen in Appendix A. For convenience and simplicity, all three assessments within the pre- and posttests were given

to the participant within a single PowerPoint presentation. Both the pre- and posttests included the same 72 MC monosyllabic words (presented in a random order for both tests) that were chosen for their relative ease of pronunciation (i.e., they did not contain any hard-to-pronounce segment or syllable structure) and because they had not been used in instruction. Each test had 6 instances of each individual tone (T1, T2, T3, and T4) for a total of 24 tokens per test, 72 total.

A researcher was present during all sessions in case participants had questions or issues related to the equipment used and the instructional materials. For consistency and to ensure that all participants completed the tests in similar circumstances, (e.g., similar timeframe and lack of repetitions), participants were not able to repeat or go back at any point during either the pre- or posttests.

1) Perception task 1: In the form of a multiple-choice question test (MCQ), the PowerPoint presented 24 MC monosyllabic vocabulary items with matching English translations and images, and the participant identified them from 4 choices representing T1, T2, T3, T4, and "I don't know". The "I don't know" option was to give the participants an option instead of randomly guessing. An example question can be seen in Figure 6.

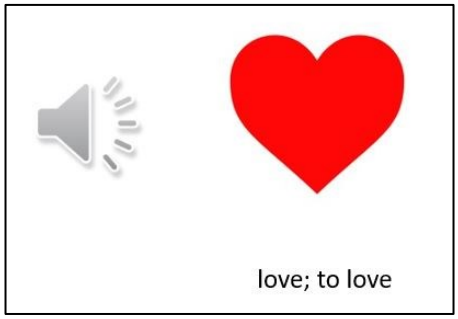

Figure 6. Example of MCQ PowerPoint slide with accompanying MC question

2) Perception task 2: The second task asks the participants to match words in an ABX test. The PowerPoint presentation presents the participant with three sounds: A, B, and X. A and B are minimal pairs with different tones, and X is a repetition of either A or B. The participant puts a checkmark beside the letter (A or B) that they believe corresponds with X. An example ABX question can be seen in Figure 7.

Figure 7. Example of ABX PowerPoint slide with accompanying ABX question.

3) Production task: Similar to the MCQ test, for the Production test, the PowerPoint presented 24 MC monosyllabic vocabulary items with matching English translations and images, and the participant attempted to repeat each MC item after hearing it. The PowerPoint slides used looked the same as the MCQ slides seen in Figure 6, but there were no accompanying written questions.

Data collection and analysis

Participants were recorded using a Microsoft LifeChat LX-3000 headset with the audio program Audacity and a portable audio recorder as a backup. Although all teaching and testing sessions were audio-recorded, only data collected in the Production test were used for the analysis of their oral performance.

To review, the research question for this study asked whether two different types of exposure would affect their perception and production to MC tones. Each MCQ and ABX question

was rated as correct or incorrect. Two native MC speakers rated the Production test results. Each production was rated as correct or incorrect by both MC speakers, and their results were compared to measure for interrater reliability. They agreed on 88% of all Production test tokens, but due to time constraints, the data from the first rater was used alone for analysis. After the perceptions and productions from the participants were rated, the findings were analyzed statistically using independent and paired-samples *t*-tests to measure variation from pretest to posttest in and across both groups.

Results

Results by test: MCQ, ABX, and Production

For the MCQ and ABX tests, items were marked as correct or incorrect by the researcher per the participants' answers on the answer sheet. For the Production test, each individual production was rated by two different raters as correct or incorrect. In the case a participant chose "I don't know", it was marked as incorrect. Both raters were native MC speakers and, across 2,976 tokens, agreed 88% of the time. Due to time constraints and the availability of the raters, only the ratings from the first rater were used for analysis. The descriptive statistics for the three tests appear in Table 1. Each tone was seen or produced 6 times in each test, and so means (M) were calculated out of 6.

To compare the groups, independent-samples t-tests were run between the T3 and T4 groups' (G3 and G4, respectively) pretests to first confirm if the groups were comparable, and then again in the posttest to test for observable differences after instruction. The three tests were the Multiple-Choice Question (MCQ) test, the ABX matching test, and the Production test. The results for the pretests can be seen in Table 1.

Table 1

Pretest data for all three tests

	G3			G4		
Pretest	*n*	*M*(/24)	*SD*	*n*	*M*(/24)	*SD*
MCQ	31	10.68	4.69	31	9.81	4.18
ABX	31	23.39	1.05	31	23.32	1.08
Production	31	20.52	3.71	31	20.42	2.35

Note. MCQ = Multiple-choice question test; ABX = matching test.

There were no significant differences between groups on the three pretests: MCQ: $t(60) = .77$, $p = .44$; ABX: $t(60) = .24$, $p = .81$) and Oral Production: $t(60) = .12$, $p = .90$. Therefore, the groups were comparable at the beginning of the treatment.

The posttest data can be seen in Table 2. To adjust for multiple tests, alpha for posttest and pretest to posttest comparisons was set at .05, and *p* values of each test were then adjusted using the False Detection Rate method (see Larson-Hall, 2010 for details on FDR). The adjusted *p* values are given for each t-test below and are only significant at $p < .05$. The posttest results, seen in Table 2, show that there were no significant differences across groups for the MCQ test ($t(60) = 1.82$, $= .14$) and ABX test ($t(60) = 1.53$; $p = .24$). However, for Production, we observe significant differences between the groups, after instruction: $t(60) = 3.49$, $p = .005$, with an effect size of .88 (Cohen's *d*) indicating a large effect.

Table 2

Postest data for all three tests

	G3			G4		
Pretest	*n*	*M*(/24)	*SD*	*n*	*M*(/24)	*SD*
MCQ	31	12.97	5.23	31	10.68	4.69
ABX	31	23.52	0.93	31	23.16	0.90
Production	31	22.19	1.72	31	20.55	1.98

Note. MCQ = Multiple-choice question test; ABX = matching test.

Within-group results

Within-group paired sample *t*-tests were run in addition to the pre- and posttest across group comparisons seen above to discover whether there were significant improvement within either group. This is to verify whether a group indeed improved from pretest to posttest, as the above data only show whether there were significant differences between the two groups.

Table 3 and Table 4 contain the same data seen in Tables 1 and 2, but reorganized to better observe within-group differences. Let us start with Table 3, which shows the within-group results for G3. This time, the participants did show significant improvement in accurate perception in the MCQ test: $t(30) = -3.742$, $p = .005$, while the ABX test did not show any change: $t(30) = -.680$, $p = .60$. G3's Production test continued to show significant improvement between the pre- and posttest; $t(30) = -3.010$, $p = .02$, but it was not quite as powerful as the between T3/T4 results reported in Tables 1 and 2.

Table 4 illustrates the within-group results for G4, wherein there were no significant within-group differences: MCQ ($t(60) = -1.129$, $p = .52$), ABX ($t(60) = -.90$, $p = .56$), and Production ($t(60) = -259$, $p = .84$).

Table 3

G3's pretest and posttest results

	MCQ			ABX Test			Production Test		
Tests	*n*	*M*(/24)	*SD*	*n*	*M*(/24)	*SD*	*n*	*M*(/24)	*SD*
Pretest	31	10.68	4.69	31	23.39	1.05	31	20.52	3.71
Posttest	31	12.97	5.23	31	23.52	0.93	31	22.19	1.72

Note. MCQ = Multiple-choice question test; ABX = matching test.

Table 4

G4's pretest and posttest results

	MCQ			ABX Test			Production Test		
Test	*n*	*M*(/24)	*SD*	*n*	*M*(/24)	*SD*	*n*	*M*(/24)	*SD*
Pretest	31	9.81	4.18	31	23.32	1.08	31	20.42	2.35
Posttest	31	10.68	4.69	31	23.16	0.90	31	20.55	1.98

Note. MCQ = Multiple-choice question test; ABX = matching test.

These results indicate that, regardless of type of tone, the group that was exposed to the third Mandarin tone (T3) during instruction improved more in perception and production of tones than the group that targeted the fourth Mandarin tone (T4).

Results by Tone

Perception. As the ABX test again did not show significant results, we present only the results from the MCQ test, which can be seen in Table 5 and 6 below. Overall, in the MCQ test, only G3 saw significant improvement, and only in the tone instructed in, T3: $t(60) = -4.429$, $p = .002$; with a very large effect size of .94 (Cohen's *d*). G3 saw no other significant changes between pre- and posttest.

Table 5.

G3 MCQ results by tone

	Pretest	Posttest	Difference
Tone	*M*(/6)	*M*(/6)	*(pre-post)*
T1	2.48	2.55	.07
T2	2.42	2.71	.29
T3	3.71	5.32	1.61
T4	2.06	2.39	.33

Note. Pretest and Posttest scores are /6

Unlike group instructed in T3, G4 saw no significant changes between pre- and posttest.

Table 6.

G4 MCQ results by tone

	Pretest	Posttest	Difference
Tone	*M*(/6)	*M*(/6)	*(pre-post)*
T1	2.65	2.32	-0.33
T2	1.9	2	0.1
T3	3.71	3.87	0.16
T4	1.55	2.48	0.93

Note. Pretest and Posttest scores are /6

One note of interest is that neither group showed much improvement in accurately perceiving T1, the least marked tone. These results indicate that exposure to a more marked tone (T3 or T4) did not have an effect on either group's perception of T1, the least marked tone. However, this was not true for the Production test, as will be described next.

Production. The production test only showed improvement with G3, as can be seen in Tables 7 and 8 below. They saw significant improvement in one tone, T1: ($t(60) = -2.800$, $p = .03$).

Table 7.

G3 Production results by tone

	Pretest	Posttest	Difference
Tone	*M*(/6)	*M*(/6)	(pre-post)
T1	4.87	5.42	.55
T2	4.97	5.42	.45
T3	5.39	5.74	.35
T4	5.29	5.61	.32

Note. Pretest and Posttest scores are /6

Table 8.

G4 Production results by tone

	Pretest	Posttest	Difference
Tone	*M*(/6)	*M*(/6)	(pre-post)
T1	4.48	5	.52
T2	5.03	4.45	-.58
T3	5.23	5.45	.22
T4	5.68	5.65	-.03

Note. Pretest and Posttest scores are /6

As illustrated in Table 8, G4 had no significant improvements in any of the targeted tones.

The main results of the study are summarized in Table 9, showing that G3 improved in terms of MCQ perception (T3 tone) and production (T1 tone) from pre- to posttest, while G4 saw no significant changes. To summarize, when these results are viewed strictly as a comparison of which group more accurately perceived and/or produced MC tones in the posttest, G3 was clearly more successful. When broken down by tone, G3 continued to be the only group to show improvement, as they saw powerful progress in their perception of T3 in the MCQ test, and in their production of T1 in the Production test.

Table 9.

Summary of results

Test-group	MCQ	ABX	Production
G3	T3 Improved	No-change	T1 Improved
G4	No-change	No-change	No-change

Note. "Improvement" indicates a significant change from pre- to posttest.

These results are explored in greater detail below.

Discussion

This study examined the effects of two types of teaching on the short-term development of MC tones: One that focused on the most marked Mandarin tone (T3) and another on the less marked fourth Mandarin tone (T4). As such, it addressed a limitation of previous studies on the acquisition of phonological developmental sequences (Cardoso, 2011a and Cardoso & Collins, 2015) by restricting the target of instruction to one single tone. This was an important issue to address because their findings favoured the group that received instruction in one single form, which also happened to be the most marked sC structure /st/, as discussed earlier. In other words, their analyses were unclear on whether the results were due to an instructional focus on the most marked /st/ form, or whether they were triggered by a mere focus on a single form by the group that overperformed in sC production. In this study, our teaching targets included a single item from the two ends of the developmental hierarchy that characterizes the acquisition of tones: T3 and T4. Accordingly, there are many similarities between the methodology adopted in this study and the one employed in Cardoso (2011a) and Cardoso and Collins (2015); however, our study involved another target language (MC), a different phonological feature (tones), and a more restrictive teaching setting, as mentioned above. In this section, we discuss the results obtained in the perception and production of tones in Mandarin Chinese.

Perception

As seen in the results section, the MCQ perception test produced significant results, whereas the ABX test did not. In the MCQ test, only the group exposed to the most marked tone (T3) showed any improvement. This improvement was seen overall between pre- and posttest, but when broken down by tone, it was only confirmed for T3, the tone in which G3 was instructed. Although these results do not seem to support, in the context of perception, Zobl's Projection of Markedness (1983, 1985), it does corroborate findings in Cardoso (2011a) and Cardoso and Collins (2015), but with a caveat. If you recall, the above authors found that knowledge of a more marked structure led to the acquisition of less marked ones. In this study, our participants only improved in T3, the most marked structure. Primarily, we believe that this might have been caused by T3's perceptual salience in comparison with the other MC tones (Maddieson, 1977; Nguyen & Macken, 2008). Consider, for instance, our results on the pretest for perception, in which T3 was identified correctly on average (but not significantly) far more than the other tones (see Table 5). Subsequently, it is possible that the focused T3 instruction that the participants received in group G3 helped them improve even further their ability to perceive the target tone. The caveat in these results is that the most perceptible tone (according to Maddieson, 1977 and Nguyen & Macken, 2008) was also the target of instruction (a factor that contributes to learning, as per Cardoso, 2011a). Clearly, our data are insufficient to answer with certainty if the observed patterns can be attributed to a single factor, or to a combination of both salience and being the target of instruction.

However, we must emphasize that the MCQ test measured *perception*, which complicates the above comparison. If you recall, Cardoso (2011a) and Cardoso and Collins (2015) found evidence that instruction of the most marked structure led to more accurate *production* of forms that comprise developmental sequences. In our study, we chose to explore both perception and

production of MC tones, for which we found that the group exposed to only the most marked MC tone during instruction (T3) was able to perceive tones more accurately than the group that received instruction in a less marked tone (T4). These results correspond with the Production test findings that saw an overall improvement, despite a difference in individual tone gains (the Production test showed improved performance in T1, not T3, as will be discussed in the next section). In other words, an *overall* improvement in perception seems to be directly correlated with an *overall* improvement in production, as observed in our investigation (see Flege, 1999 for an overview of the literature suggesting a correlation between perception and production). Intriguingly, the ABX test told a different story.

In the ABX perception test, participants were asked to decide whether X (the last sound they heard) was a replication of either A or B (i.e., the first or second sound they heard, respectively). The participants received perfect or near perfect scores so often that it was clear there was something problematic about the task. After reviewing the debriefings from the end of each participant's session, we believe that the results might be due to the task's lack of complexity. For example, the X sound (the target sound for comparison in the ABX test) was always the exact same recording as that which was used for its matching sound (A or B). This meant that participants were not actually required to compare tones as much as compare recordings. Consequently, many participants paid little or no attention to listening to the first stimulus A while performing the task, simply because they only had to decide whether B and X were the same or not. As one participant put it: "I don't even really listen to the first sound. I only listen to the second one and, if the two match, then I know it's B." It seems that it was not testing perception of tones per say, more that it was testing our participants' ability to tell whether two subsequent sounds were different. Despite the ABX's apparent ease and, consequently, dearth of significant data, we were nonetheless able

to find an increase in accurate tone perception in the MCQ test, as discussed earlier. In the next section, we explore the Production test's findings.

Production

The results of the Production test showed that exposure to the most marked tone (T3) led to significant improvement in overall accurate production, while exposure to a less marked tone (T4) did not. These results correspond with the findings in previous research (Cardoso, 2011a; Cardoso & Collins, 2015) in which it was found that instruction of the most marked phonological structure leads to more accurate production of the entire developmental hierarchy compared to other instruction methods. However, when broken down by tone, G3 only saw gains in production of T1 (the least marked tone).

The overall results indicated that one group (G3) outperformed the other (G4) in perceiving and producing MC tones. These results suggest that focused instruction on a more marked structure (and not necessarily on a single item) is more effective in the learning of non-targeted forms that encompass a developmental sequence than other methods (e.g., Cardoso, 2011; Cardoso & Collins, 2015). In general terms, these particular findings answer the research question that guided this study: which type of instruction is most effective for the initial stages of acquisition of MC tones in terms of perception and production?

However, we were also interested in how production of the different tones developed short-term in order to examine whether the observed patterns could be explained via Zobl's Projection of Markedness (1983, 1985). As described in the previous section, G3's improvement in the production of T1 is predicted by Zobl's Projection of Markedness hypothesis (1983, 1989), confirming that an instructional focus on the most marked T3 projects to the acquisition of other less marked forms (T1 > T4 > T2 > T3; Zhang, 2007, Li & Thompson, 1976; where > indicates

"acquired before" or "less marked than"). In our case, we found that 20-30 minutes of instruction of only the most marked tone in Mandarin led to significant overall improvement, which was extended to the least marked tone, T1. Zobl's hypothesis does not predict the order in which the less marked structures are acquired; our results, however, suggest that the less marked tones are learned in the order in which they are expected to be acquired in a natural setting, according to their order of markedness (T1 > T4 > T2 > T3). It is probable that, if our participants had had more than 20-30 minutes of tonal instruction, they would likely have seen more improvement in other more marked tones (e.g., T4 or T2), possibly following the order in which they are acquired in a natural setting (see, for example, Cardoso, 2011 and Cardoso & Collins, 2015, whose participants received approximately three hours of instruction to learn the target sC onsets, as discussed earlier). In sum, G3's overall results support these authors' findings and Zobl's predictions, as improvement of all four tones was significant when their results were combined in within-group tests. The same cannot be said of G4.

G4 saw no improvement from pre- to posttest, once again confirming the answer to our research question provided above. However, we did expect to see some improvement for G4, based on Zobl's Projection of Markedness hypothesis. We chose T4 as our less marked target for exposure (and not the least marked T1) because it contained a change in pitch (unlike T1), but also because it was more marked than T1. Based on Zobl's Projection of Markedness hypothesis (1983, 1985), it was anticipated that the instruction of the relatively more marked tone, T4, should lead to the acquisition of less marked tones which, in this case, is T1 (T1 > <u>T4</u> > T2 > T3). However, there was no improvement in T1 for G4, suggesting that the power of the Projection of Markedness hypothesis might be reserved for the instruction of only the most marked structure. Interestingly, all other studies that have examined the effect of instruction on the acquisition of developmental

sequences have focused on one or both ends of the hierarchy (the easy/least marked and the hardest/most marked end; e.g., Pienemann, 1989 and Zobl, 1983), or on the entire hierarchy (Cardoso, 2011). This includes research on both phonological (e.g., Cardoso, 2011; Cardoso & Collins, 2015) and morphosyntactic development (e.g., Yabuki-Soh, 2007). Therefore, we do not have evidence involving instruction of structures from the middle of a developmental sequence to corroborate this hypothesis at this time. Only future studies will be able to elucidate this conundrum.

Conclusion

This study examined the pedagogical use of two alternatives for teaching linguistic items that follow a developmental sequence, as is the case for tonal acquisition in Mandarin Chinese: one that focuses exclusively on the most marked structure (T3), or one that targets a less marked form (T4). Consistent with Cardoso's (2011a) and Cardoso and Collins's (2015) findings, one of the perception tests (MCQ) and the Production test yielded significant results that support Zobl's Projection of Markedness hypothesis for the acquisition of developmental sequences in L2 phonology. Specifically, we found that the group exposed to the most marked Mandarin tone (T3) produced similar or more accurate results across the four MC tones than the group exposed to only T4, a less marked tone.

There were a number of limitations that we plan to address in the future. The largest limitation was likely related to our testing methodology. The MCQ test did show significant results from pre- to posttest, but when broken down by tones, it only showed significance for the tone in which the participants were instructed. We believe that this problem could be improved in future studies perhaps by reducing opportunities for participants to guess their answers (e.g., by asking them to write down their answers on a form or to indicate their level of confidence). Most notably,

it was our design for the ABX test that truly needs to be improved, possibly via the addition of a certain degree of cognitive complexity. One of the participants explained: “I don’t even really listen to the first sound. I only listen to the second one and, if the two match, then I know it’s B”, suggesting another limitation, that the participants did not need to rely on their knowledge of tones to perform well in the ABX test. Other limitations that we plan to address in future research include: increasing the duration of the study to at least 4 weeks, as seen in previous studies (Cardoso, 2011; Cardoso & Collins, 2015), and to examine the power of the Project of Markedness Hypothesis to determine whether it is only applicable to the teaching of the most marked structures.

Despite these drawbacks, our findings suggest that the teaching of the most marked MC tone (T3) leads to similar or more accurate perception and production of tones than instruction of a less marked tone (T4). Pedagogically, combined with evidence from previous research involving L2 phonology (Cardoso, 2011a; Cardoso & Collins, 2015), our results suggest that the instruction of the most marked structure is the most efficient way for learners to acquire all phonological structures within a markedness hierarchy or developmental sequence. This means that teachers should emphasize aural/oral elements that are harder to acquire/more marked during instruction, and in cases where there is limited in-class time to devote to pronunciation, teachers could even forgo instruction of less marked forms as their students would likely acquire them on their own by projecting their knowledge of the most marked structures.

Chapter 3

This chapter provides a general discussion of the study presented in Chapter 2 in a broader context. Specifically, we interpret our main findings in terms of perception and production, address key questions that arose from our results, and summarize the effects of instruction observed in our study. We also discuss the directions we plan to take in the near future in terms of the acquisition of MC tones and its pedagogy, phonological acquisition, and the effects of markedness and related constructs on instruction.

Perception

As discussed in Chapter 2, the focus of this study was to examine the initial stages of L2 perception and production development, in accordance with Celce-Murcia et al.'s (2010) framework for pronunciation instruction and the acquisition of L2 phonology. For perception, we looked at listening discrimination (the first stage of pronunciation instruction as per Celce-Murcia et al., 2010) via the multiple-choice question (MCQ) and ABX tests.

In terms of perception, only the MCQ tests found significant results as they showed that, overall, G3 outperformed G4 in tone production and perception. Interestingly, when the results are broken down by tone, G3 only saw significant improvement in the perception of T3, the tone in which they were instructed. Similar findings can be observed in Cardoso (2011a), in which he investigated the acquisition of sC clusters using relatively similar instructional methods as the ones utilized in this study; e.g., while one group received instruction on the most marked /st/ structure, the other was instructed in a piecemeal fashion in the other in which /s/ + consonant forms are acquired. Although Cardoso's study did not investigate learners' perceptual development, his findings indicate that, overall, the participants performed well in the forms in which they were instructed. In our study, our participants in the more marked group (G3) saw an increase in

accurately perceiving the tone in which they were instructed, T3. However, G4 did not see any significant improvement in any tone from pre- to posttest. Therefore, although the results conform with Cardoso's (2011a) results mentioned above, they also conform with our expected results: G3 is more accurate in perceiving tones overall, and specifically T3 (a more marked tone), than G4 is in perceiving any of the tones considered in this investigation. In the following, we discussion our Production test's results in the above context.

Production

After completing the perception tests, participants performed an oral production test that asked them to mimic native MC speaker productions, thus testing their ability to produce tones in a non-phonemic/non-distinctive way. As our perception tests corresponded with the first stage of pronunciation learning/teaching, as per Celce-Murcia et al. (2010), this level of production corresponds with the second stage, controlled practice. Our results indicate that, in terms of tone production, G3 was more successful than G4 at posttest. The group exposed to the most marked tone, G3, not only showed significant improvement overall, but also in the least marked tone, T1. Looking first at the participants' production overall, we can see that the teaching of the most marked tone led to improvement in its production, supporting our hypothesis. However, unlike the perception tests that found improvement in the tone in which the participants were instructed (T3), this time, we saw an improvement in production for the tone in which the participants received no instruction, T1. Although not intuitive, these findings nonetheless support Zobl's Projection of Markedness (1983, 1985) and corroborate those found in Cardoso (2011a) and Cardoso and Collins (2015). Zobl's Projection of Markedness theory predicts that the instructional effects of targeting the most marked structure within a developmental sequence will project to the acquisition of the less marked forms. The only studies we know of that tested Zobl's hypothesis on L2

phonological acquisition found corroborating evidence that targeting the most marked structure is indeed the most effective method for instruction (i.e., Cardoso, 2011a; Cardoso & Collins, 2015). In our study, this projection effect was observed in both the transfer of T3 knowledge into the less marked T1 and in the overall improvement in tone production found in group G3. In the following, we explore G4's results, which showed no improvement.

G4 received instruction exclusively in the fourth Mandarin tone, T4. As T4 is less marked than T3 but more marked than T1 (T1 > T4 > T2 > T3), we predicted that the group would improve in the production of T4 (the target of instruction) and, consequently, T1 (the least marked tone in the hierarchy). However, this prediction was not borne out. We conjecture that this might indicate that Zobl's Projection of Markedness theory (1983, 1985) is only effective when applied to the most marked structure in a developmental sequence. Studies that have explored the instruction of developmental sequences have generally targeted multiple structures (e.g., Pienemann, 1989, 2005), only the most marked structures (e.g, Zobl, 1983, 1985), or both (e.g., Cardoso, 2011; Cardoso & Collins, 2015). We were unable to find examples where structures in the middle of a markedness hierarchy were targeted exclusively. Therefore, without evidence to the contrary, our findings suggest that only exposure to the most marked structure in a developmental hierarchy can lead to the acquisition of less marked ones. However, as our instruction time was limited (approximately 30 minutes), we cannot conclusively confirm this hypothesis, as our participants might have shown more substantial improvement if given additional time to learn. Similar to G4, the limited time devoted to instruction might also have influenced the results for G3, as these participants only saw improvement in a single tone, the least marked T1 (T1 > T4 > T2 > T3).

Despite these unexpected findings, we see that G3 saw an improvement for both the MCQ and Production tests, whereas G4 did not, thus supporting our hypothesis that the group exposed

to the more marked tone would be more successful. However, when we look at results by tone, we are left with one other key question: Why did G3's perception and production tests show specific improvement in different tones?

Perception vs. Production

The MCQ perception test showed that G3 saw significant improvement only in the tone instructed in, T3, while the Production test showed improvement in the least marked tone, T1. Either tests' findings can be viewed as predictable through the body of research around instructional methods through developmental sequences. Cardoso (2011a) found that participants often first learn the structure instructed (T3), and Zobl's Projection of Markedness (1983, 1985) predicts that instruction of the most marked tone should lead to instruction of less marked tones (e.g., T1). However, neither test showed improvement in both T1 and T3. These findings argue against the above studies. We believe a possible explanation for these patterns could be due to the nature of perception and production, which are not always acquired at the same time (e.g., Cardoso, 2011b; Flege, 1999). In Cardoso's analysis of the literature surrounding perception and production in L2 acquisition, he found examples of perception coming before production, production before perception, and both developing simultaneously. Fledge (1999) specifically reviewed the literature surrounding perception and production and found that although there is a positive correlation between the acquisition of perception and production, it is very modest. Fledge goes on to argue that differences in acquisition are likely related to the difficulty in general for L2 learners to bring production in line with their perception. Our results are commensurate with Fledge's findings as both production and perception saw overall improvement. However, although participants were able to perceive the most marked tone (T3), as it is the last (and most difficult) to be acquired in its developmental sequence, they could not transfer their knowledge to production. The successful

improvement in the production of T1, the least marked tone (and easiest to acquire), is indicative of both an increase in knowledge surrounding tonal production, but also of the inability (as of yet) to produce the more difficult and more marked tones.

In summary, our findings for both perception and production of MC tones provide evidence that instruction of the most marked tone is more effective than that of a less marked tone; however, as the above discussion detailed, these findings did not lead to improvement in all less marked tones as we had hoped to see. We have been left with more questions and ideas for future research regarding the effects of instruction on the acquisition of tones, as will be outlined below.

Future Research

There are multiple avenues that we would like to embark on in our future studies. To make our results more robust, one direction for our research is to test candidates on T1 and T2, the least marked and second most marked tones respectively. This would provide symmetry for our study, as currently we have only seen results based on instructions of T3 and T4 exclusively. In addition, this would allow us to better assess the markedness hierarchy (developmental sequence) of MC tones by more rigorously testing the current order of acquisition (T1 > T4 > T2 > T3; Zhang, 2007; Li and Thompson, 1977). Testing T1 and T2, and consequently having data on the acquisition of all four MC tones, would allow us to predict and compare the results of more comprehensive studies that address other methods of instruction such as those explored in Cardoso (2011a) and Cardoso and Collins (2015) outlined below.

As mentioned above, a direction we would like to explore involves a full replication of Cardoso (2011a) and Cardoso and Collins (2015) in a new language (i.e., Mandarin instead of an artificial language), and a new structure (tones instead of sC clusters). This would require instruction based on four approaches to teaching developmental sequences: (1) one with focused

instruction on only the most marked tone, based on Zobl's Projection of Markedness hypothesis (1983, 1985); (2) one following MC tones' developmental path, going from least to most marked (in the order they are acquired in a natural setting) based on Pienemann's Processability theory (1989, 2005); (3) the addition of a group that is instructed in all four MC tones simultaneously (based on Ammar & Lightbown's (2004) recommendation); and, (4) following the current study, we would like to include a group that will focus only on the least marked tone. With the inclusion of this last group, we aim to compare, in a more comprehensive manner, methods that focus on single versus multiple structures.

Considering these options, which aim to continue the investigation of MC tone acquisition in a laboratory setting, one final and key direction we would like to explore is to develop and incorporate targeted tone pronunciation activities in a traditional (i.e., not lab-based) MC language classroom environment. Ideally, this would be done with several groups with which we would investigate tonal acquisition longitudinally both with (as was the case in this study) and without explicit pronunciation instruction. The results would allow us to better test the theories proposed for the teaching and learning of forms that follow a developmental sequence in real world situations.

References

Ammar, A., & Lightbown, P. (2004). Teaching marked linguistic structures: More about the acquisition of relative clauses. In A. Housen & M. Pierrard (Eds.), *Investigations in instructed second language learning* (pp. 167–198). Berlin: Mouton de Gruyter.

Cardoso, W. (2011a). Teaching foreign sC onset clusters. In K. Dziubalska-Kolaczyk, M. Wrembel, & M. Kul (Eds.), *Achievements and perspectives in second language acquisition of speech* (pp. 29-40). Frankfurt: Peter Lang.

Cardoso, W. (2011b). The development of coda perception in second language phonology: A variationist perspective. *Second Language Research*, 27(4), 1-33.

Cardoso, W., & Collins, L. (2015). *Developmental Sequences in Second Language Phonology: Effects of Exposure*. Paper presented at the joint American Association for Applied Linguistics and Association Canadienne de Linguistique Appliquée Conference. Toronto, Canada.

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (2010). *Teaching pronunciation: A course book and reference guide* (2nd ed.). New York: Cambridge University Press.

Chao, Y. R. (1968). *A grammar of spoken Chinese.* Los Angeles: University of California Press.

Collins, L. (2004). The particulars on universals: A comparison of the acquisition of tense-aspect morphology among Japanese-and French-speaking learners of English. *The Canadian Modern Language Review, 61*(2), 251-274.

Delogu, F., Lampis, G., & Belardinelli, M. O. (2006). Music-to-language transfer effect: May melodic ability improve learning of tonal languages by native nontonal speakers? *Cognitive Processing*, *7*(3), 203-207.

Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning*, *27*(2), 315-330.

Flege J. E. (1999) The relation between L2 production and perception. In: Ohala J, Hasegawa Y, Ohala M, Granville D, and Bailey A (Eds.), *Proceedings of the XIV International Congress of the Phonetic Sciences: Volume 2* (pp. 1273-76). Berkeley, CA: University of California.

Hao, Y. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics, 40*(2), 269-279.

Jiang, N. (2014). Introduction. In N. Jiang (Ed.), *Advances in Chinese as a second language: Acquisition and Processing* (pp. 1-9). Newcastle: Cambridge Scholars.

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York: Routledge.

Li, C., & Thompson, S. (1976). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language, 4,* 185-99.

Keenan, E. L., & Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry, 8*, 63-99.

Nguyen, H. T., & Macken, M. A. (2008). Factors affecting the production of Vietnamese tones: A study of American learners. *Studies in Second Language Acquisition, 30*, 49-77.

Maddieson, I. (1978). Universals of tone. In J. H. Greenberg (Ed.), *Universals of human language (Vol. 2): phonology* (pp. 335-366). Stanford: Stanford University Press.

Mackey, A., & Gass, S. (2005). *Second language research: Methodology and design.* Mahwah, New Jersey: Lawrence Earlbaum.

Mackey, A., & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *Modern Language Journal, 82*(3), 338-356.

Pienemann, M. (1989). Is language teachable? Psycholinguistic experiments and hypotheses. *Applied Linguistics, 10*(1), 52-79.

Pienemann, M. (2015). An outline of Processability Theory and its relationship to other approaches to SLA. *Language Learning, 65*(1), 123-151.

Rohr, J. (2014). Training naïve learners to identify Chinese tone: An inductive approach. In N. Jiang (Ed.), *Advances in Chinese as a second language: Acquisition and processing* (pp. 158-178). Newcastle: Cambridge Scholars.

Qu, C. (2013). *Representation and acquisition of the tonal system of Mandarin Chinese* (Unpublished doctoral dissertation). McGill University, Montreal.

Lightbown, P. M. & Spada, N. (1990). Focus-on-form and corrective feedback in communicative language teaching. *Studies in Second Language Acquisition, 12*(4), 429-448.

Saito, K., & Wu, X. (2014). Communicative focus on form and second language suprasegmental learning: Teaching Cantonese learners to perceive Mandarin tones. *Studies in Second Language Acquisition*, *36*, 647-680.

Shirai, Y. (1997). Linguistic theory and research: Implications for second language teaching. In G. Tucker & D. Corson (Eds.), *The encyclopedia of language and education: Vol. 4* (pp. 1–9). Second language education. Dordrecht: Kluwer Academic.

Thornbury, S. (2002). *How to teach vocabulary.* Essex: Longman.

Triskova, H. (2017). Acquiring and teaching Chinese pronunciation. In I. Kecskes (Eds.), *Explorations into Chinese as a second language* (pp. 3-30). Albany, New York: Springer.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America, 106*(6), 3649-3658.

Westermeyer, R., & Westermeyer, J. (1977). Tonal language acquisition among Lao children. *Anthropological Linguistics, 19*(6), 260-264.

White, J., Muñoz, C., & Collins, L. (2007). The His/Her challenge: Making progress in a 'regular' L2 programme. *Language Awareness, 16*(4), 278-299.

Yabuki-Soh, N. (2007). Teaching relative clauses in Japanese: Exploring alternative types of instruction and the projection effect. *Studies in Second Language Acquisition, 29*, 219-252.

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

Zhang, H. (2007). *A phonological study of second language acquisition of Mandarin Chinese tones* (Unpublished doctoral dissertation). University of North Carolina, Chapel Hill.

Zhu, H., & Dodd, B. (2000). The phonological acquisition of Putonghua (Modern standard Chinese). *Journal of Child Language, 27*, 3-42.

Zobl, H. (1983). Markedness and the projection problem. *Language Learning, 33*, 293–313.

Zobl, H. (1985). Grammars in search of input and intake. In S. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 329-344). California: Newbury House Publishers.

Appendix A
Questionnaire and Demographic Data

Questionnaire

1. Place of birth: _____

2. Gender: _____

3. How old are you? _____
If you don't want to tell me your age, circle your approximate age?
18-25 | 26-35 | 36-45 | 45 or older

4. Education Level (e.g. high school, CEGEP, undergraduate, graduate, etc.):

5. Proficiency in **English**? Circle one: Beginner | intermediate | advanced | native

6. Proficiency in **French**? Circle one: Beginner | intermediate | advanced | native

7. What is your **first or native** language? _____

8. Do you speak or know in any other languages? YES ( ) NO ( )

If YES, please list them below and circle the options that correspond to WHERE you learned it and your overall PROFICIENCY.

LANGUAGE	**WHERE YOU LEARNED**	**PROFICIENCY (circle one)**
	Classroom | Independently	Beginner | intermediate | advanced | native
	Classroom | Independently	Beginner | intermediate | advanced | native
	Classroom | Independently	Beginner | intermediate | advanced | native

9. On a scale from 1 to 5 (1= very low, 2= low; 3= neutral; 4= high; 5= very high), how would you rate the following?

Your **musical ability**? (1) (2) (3) (4) (5)

Your ability to **imitate accents**? (1) (2) (3) (4) (5)

Appendix B
Demographics

Demographics Information

n = 62	Gender	Age Range	Education	First Language
G3	11 Male, 20 Female	20-62	6 CEGEP, 12 Undergraduate, 12 Graduate, 1 N/A	19 English, 5 French, 7 other (Gujarati, Hindi, Italian, Korean, Portuguese, Tamil).
G4	7 Male, 24 Female	20-49	10 Undergraduate, 20 Graduate, 1 N/A	15 English, 4 French, 12 other (German, Gujarati, Hungarian, Moroccan, Polish, Portuguese).

Appendix C
Instruction video scripts and screenshot

Figure A. Screenshot of an instructional video teaching the pronunciation of [ma1] ‘mother’.

Script of ten T4 MC vocabulary items

1) wò
   - wò is a third tone Mandarin vocabulary word. It means ‘grasp’ or ‘hold’.
   - Follow the guide for pronunciation: start at 2, move to 1, and then finish at 4. Like this: “wò”
   - Repeat after me 3 times: wò (wait 3 seconds), wò (wait 3 seconds), wò.
   - Now we will use the vocabulary item in an English sentence.
     - wò my hand.
     - wò onto something.
   - Now, how do you say “to hold” in Mandarin? (wait 3 seconds).
   - Make up a sentence in which you use the Mandarin word for ‘to grasp or hold’.
2) bà
   - bà is a third tone Mandarin vocabulary word. It means ‘father’.

- Follow the guide for pronunciation: start at 2, move to 1, and then finish at 4. Like this: “bà”
- Repeat after me 3 times: bà (wait 3 seconds), bà (wait 3 seconds), bà.
- Now we will use the vocabulary item in an English sentence.
    - He is my bà.
    - bà was my first word.
- Now, how do you say ‘father’ in Mandarin? (wait 3 seconds).
- Make up a sentence in which you use the Mandarin word for ‘father’.

3) dà
    - dà is a third tone Mandarin vocabulary word. It means ‘big’ or ‘large’.
    - Follow the guide for pronunciation: start at 2, move to 1, and then finish at 4. Like this: “dà”
    - Repeat after me 3 times: dà (wait 3 seconds), dà (wait 3 seconds), dà.
    - Now we will use the vocabulary item in an English sentence.
        - That building is dà.
        - I have a dà project due.
    - Now, how do you say ‘big’ in Mandarin? (wait 3 seconds).
    - Make up a sentence in which you use the Mandarin word for ‘big’.
4) zuì
    - zuì is a third tone Mandarin vocabulary word. It means ‘most, the most’.
    - Follow the guide for pronunciation: start at 2, move to 1, and then finish at 4. Like this: ‘zuì’.
    - Repeat after me 3 times: zuì (wait 3 seconds), zuì (wait 3 seconds), zuì.

- Now we will use the vocabulary item in an English sentence.
    - He is the zuì small cat.
    - She is the zuì pretty.
- Now, how do you say ‘most, the most’ in Mandarin? (wait 3 seconds).
- Make up a sentence in which you use the Mandarin word for ‘most, the most’.

5) liàn
    - liàn is a third tone Mandarin vocabulary word. It means ‘to practice’.
    - Follow the guide for pronunciation: start at 2, move to 1, and then finish at 4. Like this: ‘liàn’.
    - Repeat after me 3 times: liàn (wait 3 seconds), liàn (wait 3 seconds), liàn.
    - Now we will use the vocabulary item in an English sentence.
        - I like to liàn vocabulary.
        - I need more liàn.
    - Now, how do you say ‘to practice’ in Mandarin? (wait 3 seconds).
    - Make up a sentence in which you use the Mandarin word for ‘to practice’.
6) shòu
    - shòu is a third tone Mandarin vocabulary word. It means ‘to receive’.
    - Follow the guide for pronunciation: start at 2, move to 1, and then finish at 4. Like this: ‘shòu’.
    - Repeat after me 3 times: shòu (wait 3 seconds), shòu (wait 3 seconds), shòu.
    - Now we will use the vocabulary item in an English sentence.
        - I will shòu my package soon.
        - I have not shòu my marks.

- Now, how do you say ‘to receive’ in Mandarin? (wait 3 seconds).
- Make up a sentence in which you use the Mandarin word for ‘to receive’.

7) liàng
   - liàng is a third tone Mandarin vocabulary word. It means ‘bright’.
   - Follow the guide for pronunciation: start at 2, move to 1, and then finish at 4. Like this: ‘liàng’.
   - Repeat after me 3 times: liàng (wait 3 seconds), liàng (wait 3 seconds), liàng.
   - Now we will use the vocabulary item in an English sentence.
     - It is too liàng.
     - This room is not liàng enough.
   - Now, how do you say ‘bright’ in Mandarin? (wait 3 seconds).
   - Make up a sentence in which you use the Mandarin word for ‘bright’.
8) hèn
   - hèn is a third tone Mandarin vocabulary word. It means ‘hate, to hate’.
   - Follow the guide for pronunciation: start at 2, move to 1, and then finish at 4. Like this: ‘hèn’.
   - Repeat after me 3 times: hèn (wait 3 seconds), hèn (wait 3 seconds), hèn.
   - Now we will use the vocabulary item in an English sentence.
     - I hèn homework.
     - She hèn her class.
   - Now, how do you say ‘hate, to hate’ in Mandarin? (wait 3 seconds).
   - Make up a sentence in which you use the Mandarin word for ‘hate, to hate’.
9) qì

- qì is a third tone Mandarin vocabulary word. It means ‘air, gas’.
- Follow the guide for pronunciation: start at 2, move to 1, and then finish at 4. Like this: ‘qì’.
- Repeat after me 3 times: qì (wait 3 seconds), qì (wait 3 seconds), qì.
- Now we will use the vocabulary item in an English sentence.
    - The qì is very polluted.
    - The tank is full of qì.
- Now, how do you say ‘air, gas’ in Mandarin? (wait 3 seconds).
- Make up a sentence in which you use the Mandarin word for ‘air, gas’.

10) mài

- mài is a third tone Mandarin vocabulary word. It means ‘to sell’.
- Follow the guide for pronunciation: start at 2, move to 1, and then finish at 4. Like this: ‘mài’.
- Repeat after me 3 times: mài (wait 3 seconds), mài (wait 3 seconds), mài.
- Now we will use the vocabulary item in an English sentence.
    - She wants to mài her car.
    - What does your store mài?
- Now, how do you say ‘to sell’ in Mandarin? (wait 3 seconds).
- Make up a sentence in which you use the Mandarin word for ‘to sell’.

Script of ten T3 MC vocabulary items

1) wǒ

- wǒ is a third tone Mandarin vocabulary word. It means ‘I’, ‘me’, or ‘myself.

- Follow the guide for pronunciation: start at 2, move to 1, and then finish at 4. Like this: “wǒ”
- Repeat after me 3 times: wǒ (wait 3 seconds), wǒ (wait 3 seconds), wǒ.
- Now we will use the vocabulary item in an English sentence.
    - wǒ like to eat dumplings.
    - My boss is angry with wǒ.
- Now, how do you refer to yourself in Mandarin? (wait 3 seconds).
- Make up a sentence in which you use the Mandarin word for ‘I’.

2) bǎ

- bǎ is a third tone Mandarin vocabulary word. It means ‘to grasp’.
- Follow the guide for pronunciation: start at 2, move to 1, and then finish at 4. Like this: “bǎ”
- Repeat after me 3 times: bǎ (wait 3 seconds), bǎ (wait 3 seconds), bǎ.
- Now we will use the vocabulary item in an English sentence.
    - bǎ my hand.
    - bǎ onto something.
- Now, how do you say ‘to grasp’ something in Mandarin? (wait 3 seconds).
- Make up a sentence in which you use the Mandarin word for ‘to grasp’.

3) dǎ

- dǎ is a third tone Mandarin vocabulary word. It means ‘to hit’.
- Follow the guide for pronunciation: start at 2, move to 1, and then finish at 4. Like this: “dǎ”
- Repeat after me 3 times: dǎ (wait 3 seconds), dǎ (wait 3 seconds), dǎ.

- Now we will use the vocabulary item in an English sentence.
    - The ball dă my face.
    - Don’t dă the table.
- Now, how do you say ‘to hit’ something in Mandarin? (wait 3 seconds).
- Make up a sentence in which you use the Mandarin word for ‘to grasp’.

4) zuǐ
    - zuǐ is a third tone Mandarin vocabulary word. It means ‘mouth’.
    - Follow the guide for pronunciation: start at 2, move to 1, and then finish at 4. Like this: ‘zuǐ’.
    - Repeat after me 3 times: zuǐ (wait 3 seconds), zuǐ (wait 3 seconds), zuǐ.
    - Now we will use the vocabulary item in an English sentence.
        - You have a big zuǐ.
        - My zuǐ is dry.
    - Now, how do you say ‘mouth’ in Mandarin? (wait 3 seconds).
    - Make up a sentence in which you use the Mandarin word for ‘mouth’.
5) liăn
    - liăn is a third tone Mandarin vocabulary word. It means ‘face’.
    - Follow the guide for pronunciation: start at 2, move to 1, and then finish at 4. Like this: ‘liăn’.
    - Repeat after me 3 times: liăn (wait 3 seconds), liăn (wait 3 seconds), liăn.
    - Now we will use the vocabulary item in an English sentence.
        - A pretty liăn.
        - The ball hit my liăn.

- Now, how do you say ‘mouth’ in Mandarin? (wait 3 seconds).
- Make up a sentence in which you use the Mandarin word for ‘face’.

6) shǒu
   - shǒu is a third tone Mandarin vocabulary word. It means ‘hand’.
   - Follow the guide for pronunciation: start at 2, move to 1, and then finish at 4. Like this: ‘shǒu’.
   - Repeat after me 3 times: shǒu (wait 3 seconds), shǒu (wait 3 seconds), shǒu.
   - Now we will use the vocabulary item in an English sentence.
     - I want to shake his shǒu.
     - Use your shǒu to cover your mouth when you sneeze.
   - Now, how do you say ‘mouth’ in Mandarin? (wait 3 seconds).
   - Make up a sentence in which you use the Mandarin word for ‘hand’.
7) liǎng
   - liǎng is a third tone Mandarin vocabulary word. It means ‘two’ or ‘two of’.
   - Follow the guide for pronunciation: start at 2, move to 1, and then finish at 4. Like this: ‘liǎng’.
   - Repeat after me 3 times: liǎng (wait 3 seconds), liǎng (wait 3 seconds), liǎng.
   - Now we will use the vocabulary item in an English sentence.
     - I have liǎng feet.
     - He has had liǎng car accidents.
   - Now, how do you say ‘two’ or ‘two of’ in Mandarin? (wait 3 seconds).
   - Make up a sentence in which you use the Mandarin word for ‘two’.
8) hěn

- hěn is a third tone Mandarin vocabulary word. It means ‘very’ or ‘really’.
- Follow the guide for pronunciation: start at 2, move to 1, and then finish at 4. Like this: ‘hěn’.
- Repeat after me 3 times: hěn (wait 3 seconds), hěn (wait 3 seconds), hěn.
- Now we will use the vocabulary item in an English sentence.
    - You are hěn cool.
    - I am hěn tired.
- Now, how do you say ‘very’ in Mandarin? (wait 3 seconds).
- Make up a sentence in which you use the Mandarin word for ‘very’.

9) qǐ

- qǐ is a third tone Mandarin vocabulary word. It means ‘to get up’ or ‘raise’.
- Follow the guide for pronunciation: start at 2, move to 1, and then finish at 4. Like this: ‘qǐ’.
- Repeat after me 3 times: qǐ (wait 3 seconds), qǐ (wait 3 seconds), qǐ.
- Now we will use the vocabulary item in an English sentence.
    - Early to bed, early to qǐ.
    - qǐ your plate and clean it.
- Now, how do you say ‘to get up’ or ‘raise’ in Mandarin? (wait 3 seconds).
- Make up a sentence in which you use the Mandarin word for ‘raise’ or ‘to get up’.

10) măi

- măi is a third tone Mandarin vocabulary word. It means ‘to buy’.
- Follow the guide for pronunciation: start at 2, move to 1, and then finish at 4. Like this: ‘măi’.

- Repeat after me 3 times: măi (wait 3 seconds), măi (wait 3 seconds), măi.
- Now we will use the vocabulary item in an English sentence.
    - I want to măi a puppy.
    - I măi too much today.
- Now, how do you say ‘to buy’ in Mandarin? (wait 3 seconds).
- Make up a sentence in which you use the Mandarin word for ‘to buy’.