Beyond the walls of classrooms: Exploring the pedagogical effectiveness of text-to-speech-based shadowing on the development of Mandarin tones

Sue-Anne Richer

A Thesis in

The Department

of

Department of Education

Presented in Partial Fulfillment of the

Requirements for the Degree of Master of Art

(Applied Linguistics) at Concordia University

Montreal, Quebec, Canada

July 2023

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Sue-Anne Richer

Entitled: Beyond the walls of classrooms: Exploring the pedagogical effectiveness of text-to-speech-based shadowing on the development of Mandarin tones

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:

_____ Examiner

K. Papin (UQAM)

_____ Thesis Supervisor

W. Cardoso

Approved by _____

Sandra Martin-Chang, Graduate Program Director

_____

Pascale Sicotte, Dean

ABSTRACT

Beyond the walls of classrooms: Exploring the pedagogical effectiveness of text-to-speech-based shadowing on the development of Mandarin tones

Sue-Anne Richer

With limited classroom time (Collins & Muñoz, 2016), teachers struggle to provide personalized language input (listening activities) and opportunities for students to practice output (speaking). Text-to-speech synthesizers (TTS), also known as text readers, offer a possible solution by allowing students to interact with the computer anytime-anywhere, and at their own pace (Cardoso, 2022). As such, this technology has the potential to improve students' listening skills and provide flexible language practice (Little, 1995). Although TTS offers many benefits (e.g., immediate access to the language; Liakin et al., 2017), an unresolved issue is that the technology does not incorporate an output-inducing component (Fang, 2017). To address this issue and contribute to the field of computer-assisted pronunciation instruction, this study combines TTS with shadowing (i.e., the repetition of a word or phrases immediately after hearing it; Lambert, 1994), a technique that has been proven to be effective in developing L2 pronunciation (Foote & McDonough, 2017; Zajdler, 2020). By combining these two technologies, to which we will refer as “TTS-based shadowing training” (TTS-S henceforth), our approach provides learners with the benefits of both TTS (exposure to input) and shadowing (opportunities to practice output).

To determine the probability of success of this innovative approach, this study examined the pedagogical effectiveness of using TTS-S in a self-regulated learning environment to acquire tones #1 and #4 in Mandarin Chinese. While tone #1 (high tone) is relatively easy to acquire in comparison with other tones, tone #4 (descending tone) is considered one of the hardest to

produce (Hendry, 2023). The research was guided by the following research question: can TTS-S help L2 learners raise their sound awareness and improve their perception and production of the target Mandarin tones over six weeks? By means of pre-/post-tests (to assess effectiveness in pronunciation), ten beginner-level participants were asked to complete: (1) an awareness task in which they verbalize their metacognitive knowledge of Mandarin tones; (2) ABX tasks to assess their perception of Mandarin tones; and (3) a production task to evaluate the production of the target tones. Results indicate that the use of TTS-S did not yield significant enhancements in terms of awareness, perception, and production, possibly due to the presence of a ceiling effect in some of the measures adopted and other methodological limitations.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Dr. Walcir Cardoso, for believing in my project through this entire process. You have been guiding and supporting my research over the past year. I feel privileged to have had the opportunity to learn from you as your insightful recommendations and valuable suggestions have helped me shape my goals and aspirations. I am honored to have been your student and I would like to express my gratitude for the time you have taken to advise me on this journey.

I would also like to thank my mom who has been my rock since the beginning. You have always encouraged me to pursuit my dreams. Thank you for being a positive influence on my life.

Equally important, I would like to thank Dr. Kevin Papin for taking the time to read my proposal and for your unwavering support and guidance throughout the process. Your commitment to imparting knowledge has left a lasting impact on my research.

Finally, I want to sincerely thank all the people that took the time to be a part of my study! This would not have been possible without you. Thank you again for making this possible.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

Chapter One

After taking my first linguistic course, I knew that I wanted to learn more about languages. Unveiling the intricate web of linguistic intricacies and their influence on our cognitive processes fascinated me to no end: Languages became puzzles that I wanted to solve. So, I decided to study a variety of languages such as Algonquin, Xhosa, Tibetan, Russian, Chrau, Khmer, Yiddish and many other languages. Understanding how these different languages functioned from a linguistic perspective was like a game to me. It was refreshing to learn new languages from a perspective that was never taught to me in any second language (L2) courses. In my experience as a language learner, teachers often explained how to form sentences from the perspective of a native speaker, which did not always made sense to me, as many teachers would tend to give me the typical "you just need to memorize this" type of explanation. Linguistics provided the answers to the myriad of questions that had sparked within me. It presented an approach to L2 acquisition, seamlessly blending the rich tapestry of their diverse features, as found in the languages I have learned, into a cohesive and logical system. It helped me open my mind as I was exposed to brand new concepts both in the form of theoretical knowledge as well as in the form of variation across languages (e.g., in contrast to English, Algonquian distinguishes between animate and inanimate objects when categorizing them). In addition to this eye-opening experience, I was able to learn about phonology (including pronunciation), morphology, semantics, and many other aspects of language knowledge.

After many years studying languages, I made the decision to embark on a teaching journey, which provided me with valuable insights into the challenges students encounter within a classroom environment. I wanted to explore techniques and test different approaches that could help L2 students to learn the language they were studying. Soon, I realized that many learners

were often incapable of parsing sentences and often had trouble transmitting their thoughts verbally, even though their grammatical and reading abilities were labeled as “advanced” as per their curricula. Assuming that this could be due to the lack of opportunity to hear and speak the language in the classroom, I asked myself:

1. How can I help my students raise their *sound awareness* and develop their *listening skills* in the target language?
2. Can L2 learners improve their *oral abilities* on their own, without relying on other more fluent or native speakers of the target language?

Text-to-speech synthesizers and Shadowing: A match made in heaven?

In this thesis, to answer the first question, I propose the use of text-to-speech synthesizers (TTS), also known as text readers, as they offer a possible solution by allowing students to interact with the technology anytime-anywhere, and at their own pace (Cardoso, 2018; Cardoso, 2022). With TTS, students can benefit from multiple and varied exposures to the target language they are learning. As such, this technology has the potential to raise students’ sound awareness in the target language, leading to improved listening skills and offering flexible language practice opportunities (Little, 1995). Although TTS offers many benefits (e.g., immediate access to the language – Liakin et al., 2015; promotes learners' autonomy in their own phonological development – Moon, 2012; Liakin, Cardoso, & Liakin, 2017; enhanced, more easily perceptible input – Cardoso, 2018; 2022), an unresolved issue is that the technology does not incorporate an output-inducing component to promote speaking practice (Fang, 2017).

To address this issue (see also the second question above) and contribute to the field of computer-assisted pronunciation instruction, this study combines TTS with shadowing (i.e., the repetition of a word or phrases immediately after hearing it; Lambert, 1994), a technique that has

been proven to be effective in developing L2 pronunciation (Foote & McDonough, 2017; Zajdler, 2020). By combining these two technologies, which we will refer to as "TTS-based shadowing training" (TTS-S henceforth), our approach provides learners with the benefits of both TTS (exposure to input) and shadowing (opportunities to practice output). Given the absence of prior research on TTS-S, the outcome of merging these two technologies remains unknown. As of now, there exists no evidence or studies to shed light on the potential implications and effectiveness of this fusion.

The current study

To assess the pedagogical potential of this innovative approach, the present study examined its effectiveness within a self-regulated learning environment, specifically focusing on the acquisition of tones #1 and #4 in Mandarin Chinese. While tone #1 (high tone) is relatively less challenging to acquire compared with other tones, the accurate production of tone #4 (descending tone) is widely known as one of the most challenging to produce. Only two tones (#1 and #4) were selected to simplify the task, as the duration of the study was limited, and it included participants with no previous knowledge of Mandarin. In addition, these tones are easier for English speakers to acquire when learning Mandarin (Guo & Tao, 2008).

The research sought to answer the following research question:

(1) Can TTS-S help L2 learners raise their *sound awareness* and improve their *perception* and *production* of the target Mandarin tones over six weeks?

Using pre-/post-tests (to assess effectiveness in pronunciation), ten beginner-level participants were asked to complete: (1) a sound awareness task in which they verbalized their metacognitive knowledge of Mandarin tones after listening to a small set of words (e.g., ma-

tone#1 vs ma-tone#4); (2) ABX tasks to assess their perception of Mandarin tones; (3) a production task to evaluate the production of the target tones.

Chapter 2 delves into the motivation behind the research, outlines the methodology employed, presents the obtained results, and discusses the implications for the self-regulated learning of Mandarin Chinese tones. As per the guidelines for manuscript-based MA thesis, the next section constitutes “a full submittable draft of a manuscript” that presents a literature review, methodology, results, and discussion of the aforementioned research.

Chapter Two

Many Mandarin Chinese learners claim to be tone aware (Yang & Medwell, 2019). However, studies have shown that tones are difficult to acquire, especially for beginner learners (Lin, 1985; Wang et al., 2003) due to the challenges associated with recognizing and discriminating them (Hao, 2012). Acquiring tones in Mandarin is essential because tones serve to distinguish word meanings (Zhang & Gao, 2022). To help learners develop their phonological awareness of tones, as well as their perception and oral production, this study proposed to use text-to-speech synthesizers (TTS) due to the possibility of enhancing students' listening and speaking abilities by engaging them in interactions with the computer (Cardoso, 2018; Cardoso, 2022).

Although text-to-speech synthesizers offer many advantages, an unresolved issue is that they do not incorporate a production component (Fang, 2017). In order to address this issue and contribute to the field of computer-assisted pronunciation instruction, this study proposed to combine TTS with shadowing to allow students to improve their awareness, perception, and production of tones in Mandarin. It has been demonstrated that shadowing is effective in developing second language pronunciation (L2) (Zajdler, 2020; Foote & McDonough, 2017). Combining these two technologies, learners were not only exposed to Mandarin, but were also given the opportunity to produce spoken output. This "new" method of teaching will be referred to as TTS-based shadowing training (TTS-S, as will be described in forthcoming sections).

Another issue that negatively affects the L2 learning experience is the increased workload experienced by language teachers (Timperley & Robinson, 2000). Teachers simply do not have enough time to provide their students with personalized and relevant L2 input and output opportunities, particularly when they also have to cover a number of topics mandated by the

curriculum. Adopting a more sustainable model of teaching is important as it could help support the teacher to refocus on other important parts of teaching, such as providing feedback or engaging students in communicative activities.

Following Cardoso's four-stage chronological framework (2022) for investigating technological tools for L2 pedagogy and considering that the two adopted technologies already exist (stage 1: software development) and have been explored from a pedagogical standpoint (stage 2; see literature review), this study focused on stage 3: Assessing pedagogical effectiveness. This assessment was conducted using a mixed-method approach and a pretest-posttest design, considering the three targeted levels of phonological development: phonological awareness, aural perception, and oral production.

Literature Review

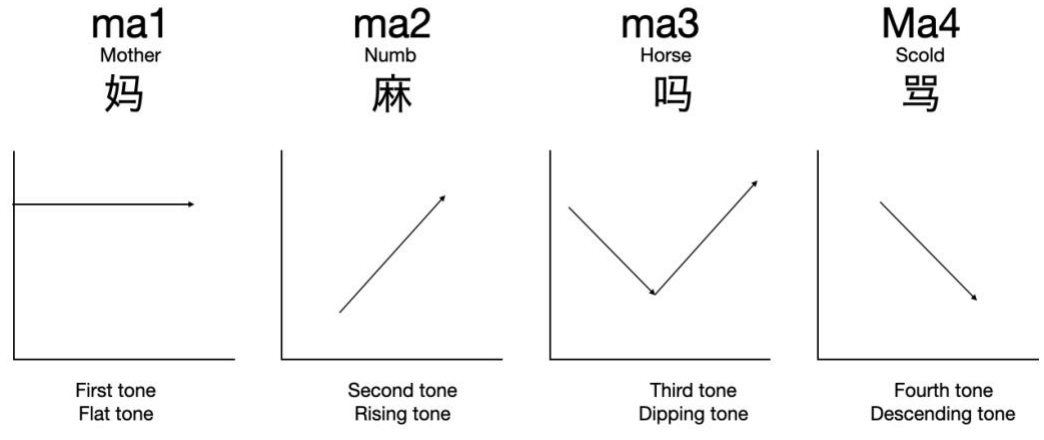
Tonal language

Mandarin

Mandarin includes four tones and a neutral tone (Yang, 2016). The number associated with the vowel in each syllable (1=high tone, 2=rising tone, 3=dipping tone, 4=descending tone, 5= no tone or neutral tone) represents the tone, as presented in Figure 1.

Figure 1

Tones



Contrary to non-tonal languages, in tonal languages such as Mandarin, the modulation of pitch (also known as tones) conveys a difference in meaning (Li et al., 2021). This suggests that the production and phonological perception of the correct tone is essential to the intelligibility and comprehensibility of the speaker (Chen et al., 2008). For instance, the numeral "nine" (几) "ji3" is produced with a dipping tone (falling-rising tone), while the homophonous word for "chicken" (鸡) "ji1" is recognized and understood by the production of a flat high tone.

Given that these distinctive tonal features play an important role for the comprehension of the language and tones are difficult to perceive and produce by learners of non-tonal languages (Triskova, 2017), practice in both perception and production of tones is necessary. Yet, tones are rarely clearly taught in traditional Chinese language courses in the initial stage of their learning experience, learners are expected to pick up the correct pronunciation from the exposure to the language (Orton & Scrimgeour, 2019). Although Mandarin Chinese learners claim to be tone aware (Yang & Medwell, 2019), studies have shown that they are difficult to acquire, especially for beginner learners (Lin, 1985; Wang et al., 2003) due to the challenges associated with recognizing and discriminating them (Hao, 2012). Acquiring tones in Mandarin is essential because tones serve to distinguish word meanings (Zhang & Gao, 2022). To help learners develop their phonological awareness of tones, as well as their perception and oral production, this study proposed to use text-to-speech synthesizers (TTS) in a self-directed learning environment due to the possibility of enhancing students' listening and speaking abilities by engaging them in interactions with the computer (Cardoso, 2018; Cardoso, 2022) which could help mitigate some of the time limitations that affect the L2 classroom. As the duration of this study is limited, and it includes participants with no previous knowledge of Mandarin, only tones

#1 and #4 were chosen to simplify the task (Guo & Tao, 2008). In addition, these tones are easier for English speakers to acquire when learning Mandarin (Guo & Tao, 2008).

Self-Directed Learning

As highlighted earlier, L2 teachers frequently face challenges in delivering individualized language input (i.e., activities that focus on the development of sound awareness and listening skills) and providing opportunities for students to practice their speaking skills within the constraints of limited classroom time (Collins & Muñoz, 2016). A potential solution to this situation could be via the empowerment of students with self-directed learning "tools" that foster independent language development. Learners who engage in self-directed learning (SLD) are responsible for their own learning while assessing their progress (Little, 1995). This implies that learners can study at their convenience according to their schedule anywhere they want, without any time constraints. In addition, self-directed learning instills lifelong learning practices (Du, 2013). Building on these insights, this study proposed to nurture lifelong learning practices in a self-directed learning environment using text-to-speech synthesis.

Text-to-speech

Text-to-speech (TTS), also known as text readers, is a tool used for converting digital text into equivalent audio output (Fang, 2017). They are found in modern applications and devices such as GPSs, voice assistants, and translation applications (Cardoso, 2022). Converting text into speech occurs as follows: the user inputs text into the TTS application, the text is then converted to a phonemic representation of the language, which is then matched to speech units that create the oral output (Jurafsky & Martin, 2000).

Modern TTS applications offer enhanced L2 exposure to the learners (e.g., speech can be manipulated so that it can be slowed; Liakin et al., 2015) and provide ample opportunities for

learners to study wherever and whenever they wish. In other words, learners can be exposed to the L2 input within or outside of the classroom, and without living in a foreign community where the target language is spoken. Another benefit to be noted is that modern synthesizers produce output of excellent quality when compared with that of human speakers (Craig & Schroeder, 2017). For instance, Bione and Cardoso (2020) found that the performance of both English synthesized voices and English human voices were similarly comprehensible, with only naturalness being favored in human voices.

A key element to consider when selecting the right TTS synthesizer is its ability to produce accurate speech. In terms of tone production, Fang (2017) found that Nuance and is the most accurate software (60.8% of the target words — 14 accurate tones out of 23) the is free and easy to use. Note that in this study, other software such as Ispeech were labeled as more accurate however as money could be a limiting factor for many users.

When learners use TTS for listening (aural) practice, it is important that they also have opportunities to improve their *oral* production skills in the target language. To promote this type of practice (oral production), this study proposed the use of shadowing, as will be discussed next.

Shadowing

A key component of becoming proficient in a second language is the comprehension of the input (Krashen, 1982). As many other researchers have shown (Ellis & Collins, 2009; Yanguas, 2012), the recognition and the perception of specific linguistic features are a prerequisite to learn a language successfully. Although being able to perceive sounds is a necessity, Swain (2005) asserted that producing output also plays an important part in learning a language. To promote the learning of Mandarin tones, this study included an oral production component: shadowing.

Shadowing is described as ‘repeating what one hears simultaneously as accurately as possible (Hamada, 2018). Shadowing uses a “top-down” approach that encourages learners to enact the voices of others (Pennington, 2021). It can be applied at different language levels and can be adjusted according to the aims set for the particular skill training (Zajdler, 2020). This technique is usually aimed at low proficiency learners, as it helps them activate the phonological perception and imitate unfamiliar speech sounds. In other words, shadowing, on a cognitive level, causes learners to pay more attention to the sounds they hear rather than the meaning of the sentence (Hamada, 2018) due to their inability to process what they hear rapidly enough (Shiki et al., 2010).

One of the numerous advantages of shadowing is that it leads to notable gains in comprehensibility (medium-to-large effects) and results in a reduction in foreign accentedness (medium effects) (Hamada, 2018). In addition to the aforementioned advantages, shadowing activities may also give learners ample opportunities to notice differences between their pronunciation and the target pronunciation (Foote & McDonough, 2017).

Although many studies have shown that shadowing helps with awareness, perception or production (Foote & McDonough, 2017; Zajdler, 2020;), one of the main obstacles of such technology is that it relies on either pre-existing recordings or humans (for real-time activities). As such, it is very limited to what the recording or the teacher can provide for the students. Fortunately, shadowing technologies can be used in conjunction with TTS software to combine the benefits of both technologies. To take a first step towards exploring TTS-based shadowing training (TTS-S, a term coined for this study) as a pedagogical tool, this study assessed its pedagogical effectiveness.

Second Language Phonological Development: Assessing Pedagogical Effectiveness

To assess pedagogical effectiveness, this study adopted Celce-Murcia's framework (2010) for teaching pronunciation, which predicts that second language phonological development progresses in stages ranging from phonological awareness through perception or aural discrimination all the way to production practice. This framework assumed a perception-precedes-production approach (Cardoso, 2011; Schmidt, 2001; Shao et al., 2022; Kachlicka et al., 2019; Mueller et al., 2012), which holds that awareness on the part of learners is an essential prerequisite for the perception of sounds, which is in turn important for the development of oral production skills (Celce-Murcia et al., 2010).

To capture the effectiveness of TTS-S, this study followed a pretest and posttest design to assess the participants' development of Mandarin tones in phonological awareness, aural perception, and oral production.

Current Study, Goals and Hypotheses

The purpose of this study was to examine the pedagogical effectiveness of TTS-based shadowing (TTS-S) for the learning of L2 tones. TTS-S has been proposed for at least three reasons. Firstly, it promotes a self-directed learning environment that permits students to study anytime anywhere without the assistance of a native or fluent speaker of the target language or a language teacher (Thornton, 2010). Secondly, TTS-S is free and easy to use, which makes it accessible to a large audience. Lastly, people can use any material of their interests, which plays a decisive role in motivating them to learn the target language, particularly in self-regulated learning environments (Liu, 2015; Ma & Ma, 2012; Gilakjani et al., 2012).

To assess the pedagogical effectiveness of TTS-S, this study addressed the first three stages of pronunciation development in L2 acquisition, as per Celce-Murcia et al. (2010): (1)

phonological awareness, (2) aural perception, and (3) controlled oral production. These three stages are important because they represent the initial developmental stages of the acquisition of an L2 phonological system (Celce-Murcia et al., 2010). Successful learners must be aware of the target feature before they are able to perceive and subsequently produce it (Baese-Berk, 2019; Shao et al., 2022). As such, this study assumed that phonological perception precedes production in the acquisition of new phonological systems (e.g., for the rationale, see Cardoso, 2011; Shao et al., 2022; Kachlicka et al., 2019; Mueller et al., 2012).

To explore these three stages of phonological development, the following research questions were addressed in this study:

1) Will the use of TTS-based shadowing in a self-directed learning aid the learning of the target Mandarin tones in terms of:
    1) Awareness?
    2) Perception?
    3) Production?

Methodology

Participants

Ten adult beginner learners of Mandarin participated in this study. The participants had no knowledge of the Mandarin tones and had never taken any Chinese courses. They came from varied non-tonal languages (e.g., English, French, Arabic), and educational backgrounds (commercial art, accounting, social studies). The participants were recruited through various channels (e.g., Facebook, Instagram). Although most of them are Canadians living in Montreal, Ottawa, or Toronto, some of the participants come from various countries (e.g., United States, Lebanon).

Learning Material

Participants were given a basic Mandarin tones introduction before the learning activities started. For instance, they were told that tones are important in Mandarin because they contribute to meaning (e.g., they serve to differentiate words), and that they would discover what these tones are. Then, they were provided with five texts containing 5 to 9 sentences in which the two targeted tones appeared a total of 66 times (tone #1 and #4 appear between 10 to 16 times a week). These texts were copied and pasted into the text-to-speech synthesizer every week (Week 1 to 5). An example of the text is available in Appendix A. If learners decided to use additional material, they were asked to add it to their weekly learning logs.

Instruments

Demographic Questionnaire

A demographic questionnaire (Appendix B), following Van Lieshout and Cardoso (2022), was used to collect the participants' background information. This questionnaire included information such as the languages they speak, whether they had learned a language in a self-directed manner if they had used iSpeech previously (the TTS application adopted for this study, as stated earlier), and their educational background.

Pretests and Posttests

To evaluate the phonological development of tones, this study looked at the effects of the treatment on the awareness, perception, and production of tones in a self-directed TTS-S based environment.

Awareness test. An awareness test (Appendix C) was used to measure the metacognitive knowledge of the students about tones in Mandarin. Participants had to complete a pre-awareness test and a post-awareness test at the beginning and end of the experiment. The tests were the

same, but with randomized items to mitigate testing effects. In the awareness test, a small set of six minimal pairs (e.g., ma1 vs ma4) were played and the participants had to label the sounds as the same or different and then explain, in writing, how these sounds are different. There were six questions in which three compared different tones (e.g., tone 1 vs. tone 4) and three compared the same tone (e.g., tone 4 vs. tone 4).

Perception tests. As the learners in this study were people with no exposure to Mandarin, it was important to present a task that was easily understood by all. To investigate the perception of the learners, this study used two ABX tests. The ABX tests were aural discrimination tests. ABX tests were selected because they are regarded as easy to administer, simple, and intuitive (Hautus & Meng, 2002; Huang & Lawless, 1998).

In the first ABX tests, participants were asked to determine whether there was an audible difference between two audio signals (Huang & Lawless, 1998). In the second ABX test, participants followed a match-to-sample discrimination procedure. This test involved the presentation of two sounds (test and control items) followed by the presentation of a third sound (blind sample) that corresponded to the test sound. The participants were asked to match the blind sample to the test item.

ABX 1. A perception test was used to assess the phonological perception (one's ability to recognize and categorize sounds) of Mandarin tones in Chinese. Both pretest and posttest for the ABX 1 test (Appendix D) were the same to ensure that the comparisons are reliable and to avoid having unbalanced data. As mentioned earlier, participants were presented with two spoken tokens and were asked if the two sounds are the same or different. An answer sheet (Appendix F) containing both ABX tests was given to the participants. The ABX 1 test, consisting of pre-recorded audio of a human voice containing 20 tokens, focusing on five different syllables (e.g.,

[ma, me, mi, mo, mu]) and the two target tones in Mandarin (e.g., tone 1 — flat tone and tone 4 — descending tone), as they are easier to acquire for English speakers (Guo & Tao, 2008). Each vowel appeared twice with the same tone for all tones. For instance, the syllable [ma] appeared twice with the first tone and twice with the fourth tone for a total of four appearances (i.e., one subset containing the vowel "a" would consist of the following two words or pseudo words: ma1, ma1, ma4, ma4). The number of appearances of each tone and related vowels was equally distributed among the tokens. A female native speaker of Mandarin audio-recorded the tokens in the test.

ABX 2. A second perception test was used to further support the results obtained in ABX1. Again, pretest/posttest 2 (Appendix E) were the same. In this ABX 2 test, participants were first presented with two sounds (control sound and test sound), then they were presented with a third sound (blind sound). They then were asked to match the (third) blind sound to the control sound. There were 20 blind sounds (10 syllables associated with each tone, tones 1 and 4) to ensure equal distribution of the syllables and their associated tones. Answers were collected on the same answer sheet (Appendix F) provided by the researcher in the first ABX test.

Production Test. To examines the production of the Mandarin tone 1 and 4, the production test (Appendix G) adopted a 'listen and repeat' activity, a less cognitively demanding task (e.g., in comparison with spontaneous speech) because of the low proficiency of the participants and the duration of the treatment (for the rationale, see Cardoso et al, 2021). Once again, this task was measured in pre-/posttests in which the learners were asked to listen to a recording of a native speaker of Mandarin and tried to repeat what they hear as accurately as possible. Everything was recorded to be further analyzed at the beginning and the end of the

experiment. There was a total of 10 tokens. Each syllable/word were associated with either tone 1 or tone 4 to ensure that the data are balanced.

Procedures and Design

Prior to engaging in the study, participants were asked to sign an informed consent form and fill out a demographic questionnaire. Participants then completed the awareness test, the two ABX tests, and the production task. The participants were then provided with a brief description and analysis of the tones in Mandarin and a short tutorial on how Ispeech, the selected text-to-speech synthesizer, works. Participants were also provided five texts (Appendix A) to use with text-to-speech-based shadowing over the following weeks. They were instructed to go over the pre-selected texts at their own pace. As per the study's recommendation, it was expected that they completed one text every week. A weekly reminder was sent to the participants to indirectly track if participants complete the activities as instructed. At the end of the experiment, all posttests were once again administered.

Figure 2 illustrates the research design and procedures adopted.

Figure 2

Design of the study

Week 1:	Demographic questionnaire Awareness Test ABX 1 ABX 2 Production task Description & Analysis Text-to-speech-based Shadowing - Text 1
Week 2:	Text-to-speech-based Shadowing - Text 2
Week 3:	Text-to-speech-based Shadowing - Text 3
Week 4:	Text-to-speech-based Shadowing - Text 4
Week 5:	Text-to-speech-based Shadowing - Text 5
Week 6:	Awareness Test ABX 1 ABX 2 Production task

Data Analysis

Demographic questionnaire

The demographic questionnaire was used to support and further develop the results obtained via the pretests and posttests.

Awareness Tests

The phonological awareness data were analyzed via a 3-level system: 0 (fully incorrect – not aware); 1 (partially correct – partial knowledge); 2 (fully correct – full knowledge of tones). If the participants could not tell the difference between the two tones they heard, they did not provide a justification of their answers, or their answer did not describe how the sounds were different, no points were granted to the participant. If the participant could either tell whether the two sounds were different or explained how they were different, they were given 1 point. If the participants were able to do both, they were given 2 points. There were six tokens in which three syllables were the same and three were different.

ABX Tests: ABX 1 and ABX 2 and Production task

Both ABX tests as well as the production task were analyzed using descriptive statistics via SPSS. The computation of means, standard deviations, and other relevant statistical analyses and significance were conducted to discover if the use of the proposed technology (TTS-S) helped beginner learners of Chinese aurally discriminate and orally produce the two target Chinese tones.

Results

Means of pretests and posttests were compared using paired sample t-tests. Participants' awareness, perception, and production of tones were tallied to determine if the utilization of TTS-S within a self-directed learning context promotes the learning of target Mandarin tones in terms of awareness, perception, and production abilities. The descriptive statistics indicated that the average scores from all four posttests were higher compared to the scores obtained in the pretests; however, these differences were not statistically significant. Table 1 provides a summary of the results for the four tests administered: sound awareness (awareness), ABX1 and ABX2 (perception), and oral production (production).

Table 1

Summary of the results

Tests	**Test (N=10)**	**Mean**	**SD**	**Average Gain Score**	**SD Gain Score**
Awareness	Pretest	44.17	10.43	5	10.54
	Posttest	49.17	2.64		
ABX1	Pretest	97	4.83	2	4.21
	Posttest	99	3.16		
ABX2	Pretest	88	16.36	6.5	11.56
	Posttest	94.50	8.31		
Production	Pretest	79.75	15.47	2.25	5.83
	Posttest	82	15.47		

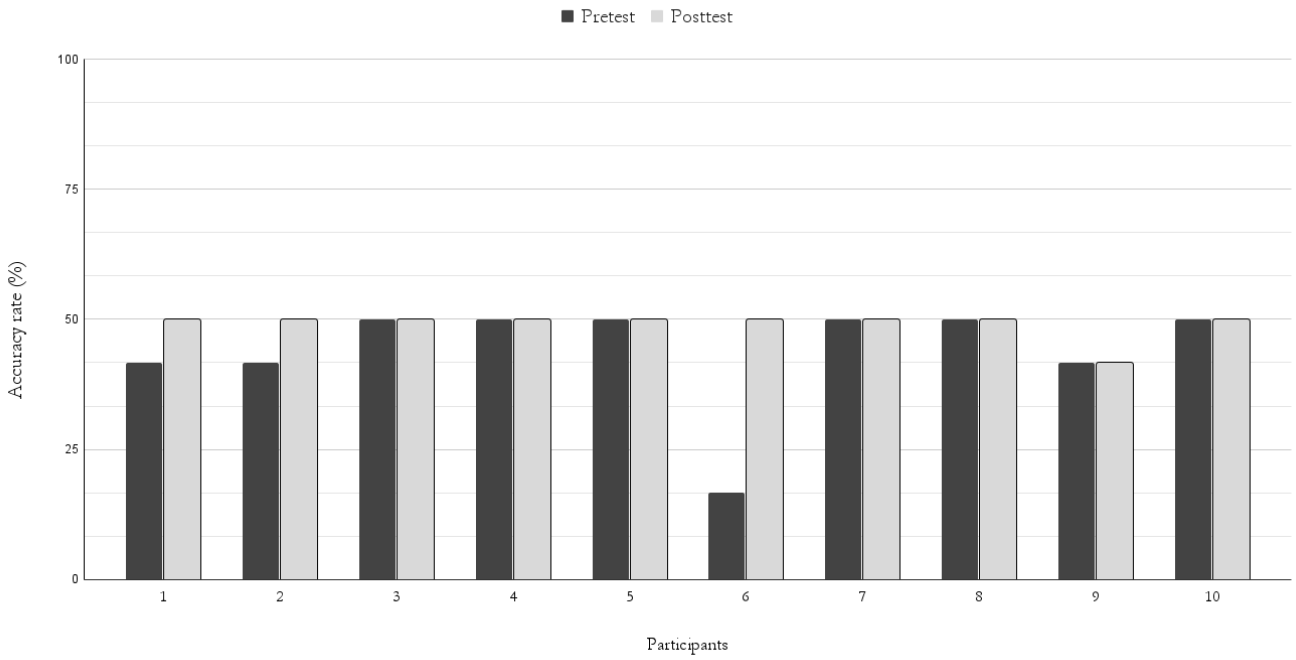
Awareness Tests

To determine whether the participants' awareness of tones would alter between the pretest and posttest, a paired t-test was performed. Results were not statistically significant ($t = 1.50$, $p = .168$, $\alpha = .05$).

Figure 3 illustrates the results for each participant. These results suggest that 3 participants (1, 2, and 6) improved their awareness over six weeks whilst the score of the other participants remained the same. An analysis of the participants' responses indicated that they were successful in identifying whether the target sounds were similar or dissimilar. However, none of the participants were able to accurately describe the underlying differences in tones in the second part of the inquiry, which specifically asked them to articulate the difference in tone using their own words. Most respondents believed that the changes in sounds were caused by the length of the vowel (ma vs. ma:), as well as its sharpness and loudness. Some explanations were not relevant (e.g., the tone in question was "more decisive" or "more inquisitive").

Figure 3

Awareness Tests: Accuracy rate (%) by participant



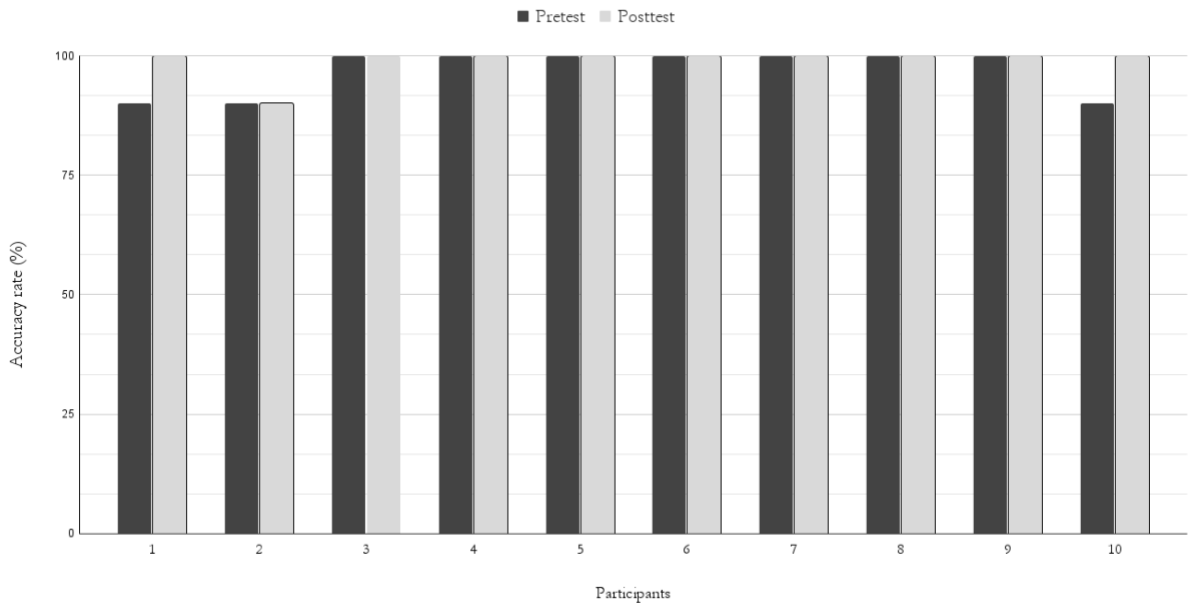
ABX Tests

ABX 1

The paired t-test indicated that the results were not significant between the pre- and posttest ($t = 1.50$, $p = .168$, $\alpha = .05$. Figure 4 displays the results for each participant, where the accuracy rate in the ABX1 test only improved for two individuals (1 and 10).

Figure 4

ABX1: Accuracy rate (%) by participant

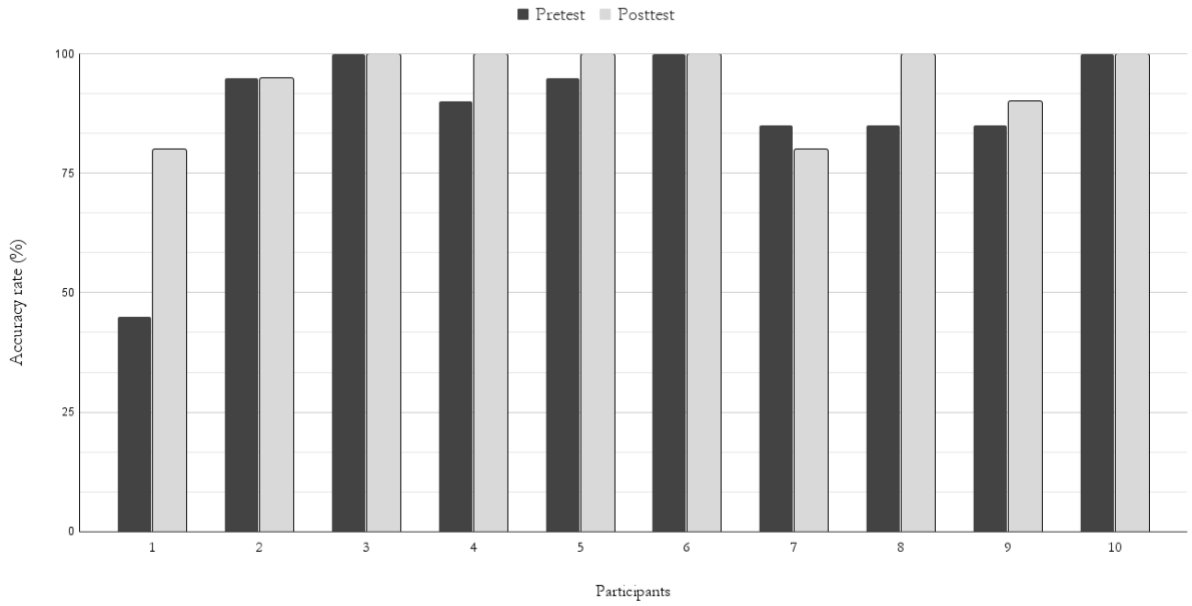


AXB2

The findings of the paired t-test showed that there was no significant difference between the pre- and posttest ($t = 1.78$, $p = .109$, $\alpha = .05$). Figure 5 illustrates the results for each participant, where it is shown that 50% of the participants (n=5) improved in the ABX2 posttest (1, 4, 5, 8, and 9).

Figure 5

ABX2: Accuracy rate (%) by participant

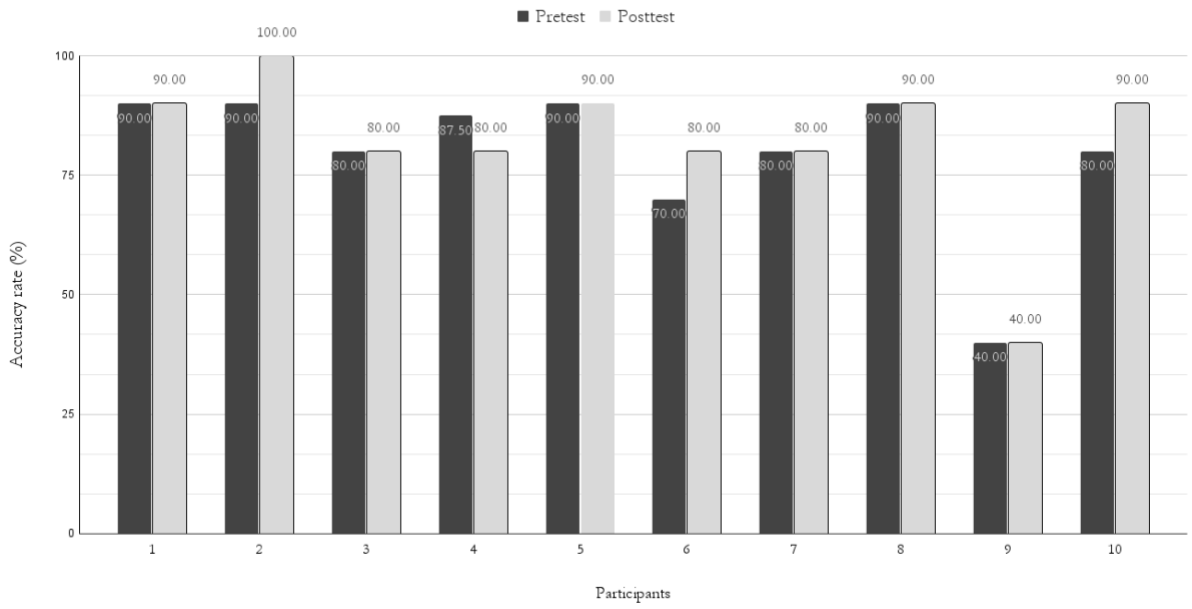


Production Task

The t-test analysis conducted on the oral production data suggest that there was no statistically significant difference between the pre-test and post-test means ($t = 1.22$, $p = 0.25$, $\alpha = .05$). As seen in Figure 6, at an individual level, it was evident that three participants (2, 6 and 10) demonstrated improvement in their pronunciation at the end of the six-week experiment.

Figure 6

Production Task: Accuracy rate (%) by participant



In summary, no statistical significance was observed in the outcomes of the four tests, namely the awareness, ABX1, ABX2, and production tests. It is important to note, however, that although some participants demonstrated improvement in tone performance, there were others who maintained their performance levels unchanged throughout the duration of the study.

Discussion

The objective of this study was to investigate the pedagogical effectiveness of a novel approach for self-regulated learning that combines TTS with shadowing for the learning of tones #1 and #4 in Mandarin Chinese. As such, it was guided by the following research question: can TTS-S help L2 learners raise their sound awareness and improve their perception and production of the target Mandarin tones over a period of six weeks? Results showed that all four tests (i.e., sound awareness, perception (ABX1 and ABX2), and production) were not statistically significant, suggesting that, overall, the proposed treatment did not yield any significant improvements in the assessed measures of pronunciation development.

In the initial section of the awareness test, participants were required to assess the similarity of the tones they heard in monosyllabic words. Out of the ten test takers, six achieved perfect scores, while three had only a single error, indicating a clear ceiling performance. The term "ceiling effect" refers to a phenomenon in which the potential improvement in test scores is limited because a participant's initial score is already positioned towards the upper end of the score distribution (Koedel & Betts, 2010). Considering that most of the participants had already reached their maximum potential for improvement, their capacity for further progress was limited.

Interestingly, the opposite was observed in the second portion of the test, in which participants were asked to describe the differences between the tones they heard. In this section of the awareness test, we observed a “floor effect” (Tavakoli, 2013), a phenomenon where the extent of improvement in test scores is constrained because the participants’ initial scores are situated at the lower end of the score distribution. In this test, for instance, participants failed to articulate the various tonal variations that were present in the words they heard, attributing

differences in tones to the length or loudness of the vowels. These patterns reflect previous research that claim that L2 learners rely on psychoacoustic cues such as loudness and the strength of sound fluctuation (e.g., vowel length differences) when discriminating between tone pairs (e.g., Yang, 2021), as was observed in the current study. Our findings thus suggest that the measure used lacks sensitivity in detecting performance variations among individuals who are performing at a low level of proficiency. This observation aligns with Pienemann's (1998) Processability Theory, which claims that L2 learners might not be capable of processing certain "advanced" learning material because they are not at the appropriate stage of language development required for effective learning. In other words, it is possible that our participants' current level of language proficiency in Mandarin Chinese may hinder their ability to fully grasp and benefit from the instructional material provided.

As for the first perception tests (ABX1), in which participants were presented with two spoken tokens and were asked whether they sounded the same or different, seven participants achieved perfect scores, while three participants achieved a score of 90%, indicating a high level of performance among all participants. The t-test results confirmed the absence of statistical significance. Again, as was the case in the first awareness test, these outcomes align with the presence of a ceiling effect (Koedel & Betts, 2010), wherein participants' initial high scores limited their potential for further improvement. The near-perfect scores obtained in this test align with previous findings by Shen and Froud (2018) and Wu and Lin (2008), which indicate that English speakers exhibit heightened perceptiveness to the contrast between Mandarin tones #1 and #4. This heightened sensitivity may be attributed to the fact that English speakers are particularly attuned to the cues present in tone 4, enabling them to differentiate it from its counterpart, tone 1 (Xu & Xu, 2005).

In the second perception test (ABX2), which asked the participants to match what they heard last (“blind sound”) with a control sound, while five people improved their scores, four people’s scores remained the same, and one person’s scores decreased. The t-test results confirmed the absence of statistical significance, supporting the notion that no significant improvements could be observed between the pretest and posttest. Once again, this finding can be attributed to a near-ceiling effect, where the performance of nine participants fell within the range of 80% to 100%. As participants were already performing at a high level at the pretest stage, there was limited room for measurable improvement in the posttest scores. These findings are in line with earlier research conducted by Shen and Froud (2018), Wu and Lin (2008), and Xu and Xu (2005), which demonstrate a heightened perception towards the Mandarin tone contrast, particularly when it comes to differentiating between tones #1 and #4.

Finally, the t-test results for oral production, which adopted a ‘listen and repeat’ test, indicated that the observed increase in scores between the pretest and posttest was not statistically significant. This lack of significance is not surprising, and it can likely be attributed to the choice of the test itself, as it represents a less cognitively demanding task compared to read-aloud or spontaneous speech. This methodological decision was made based on Cardoso et al.'s (2021) study, considering that the participants were “real” beginners in Mandarin Chinese and, as such, could not read aloud or engage in spontaneous oral interactions. Out of the ten participants, one individual's score decreased, six participants' scores remained the same, and three participants' scores increased. The wider score fluctuations observed among the participants suggest that other factors may have influenced the results. Individual factors in language learning, such as age, motivation, and learning strategies (DeKeyser, 2013; Robinson, 2010), could have contributed to the variations in participants' performance. For instance,

hypothetically, participants who lacked intrinsic motivation to learn the target tones, perceiving them as externally imposed tasks, may have been less engaged and invested in the learning process. Their limited motivation might have resulted in lower effort and dedication towards mastering the tones, potentially leading to less improvement in their scores between the pretest and posttest. Conversely, participants who possessed intrinsic motivation, driven by their personal interest, curiosity, or enjoyment of learning languages, might have been more inclined to actively engage in the learning activities. Their inherent motivation could have fueled their persistence, enthusiasm, and willingness to invest additional effort into practicing the tones, thereby facilitating their learning, and potentially resulting in more noticeable improvements in their scores from the pretest to the posttest.

Two important findings can be derived from our study. Firstly, our findings support the established notion that L2 phonology follows a trajectory that begins with the development of phonological awareness and aural discrimination (perception) and progresses towards the acquisition of oral production skills (Celce-Murcia et al., 2010; Schmidt, 2001). In this study, we observed that participants exhibited a high level of awareness and perception regarding tone 1 and tone 4, suggesting their near- or complete development. However, there remained room for improvement in the production of these tones, highlighting the sequential nature of tonal development where perception typically precedes production. Secondly, our study provides evidence that although certain aspects of tone development can be achieved within a semi-directed learning environment, as the one explored in this study via a combined use of TTS and Shadowing, the effectiveness of such interventions may vary among individuals due to inherent individual differences. Some participants may derive greater benefits from this intervention,

while others may not exhibit the same degree of improvement, highlighting the influence of personal factors on the learning process.

Conclusion

Can TTS-S help L2 learners raise their sound awareness and improve their perception and production of the target Mandarin tones over six weeks? Based on the findings of this study, it is inconclusive to assert that TTS-S can effectively raise sound awareness or significantly improve the perception and production of the target two Mandarin tones among L2 learners within a six-week timeframe. The results indicate that while certain aspects of phonological development, specifically production, showed some improvements, the overall impact was not statistically significant. The interpretation of these findings may be partially explained by the faulty methodology used or, conversely, by the participants' individual differences. Factors such as motivation and prior linguistic background may have influenced the outcomes. It is possible that some participants were more responsive to the TTS-S approach and demonstrated improvement, while others may not have benefited as significantly.

Furthermore, it is worth acknowledging the limitations of this study, which could also have impacted the results. One of the most important limitations is its small sample size of only 10 participants. Although this was done to ensure an in-depth examination of the participants' pedagogical experience (including learner perceptions, use of strategies in TTS-S/human interactions – not reported in this thesis), larger samples are needed to obtain more accurate estimates and enhance the generalizability of the findings. A methodological limitation includes the simplicity of most tests (e.g., the first awareness test, ABX1, oral production), which were found to be extremely easy, with participants achieving ceiling or near-ceiling scores. Another limitation regards the lack of an account for individual differences among the participants, such

as age, motivation, and learning strategies, which have been shown to significantly impact language learning (DeKeyser, 2013; Robinson, 2010). Further investigations are necessary to validate the pedagogical effectiveness of TTS-S and its suitability in language learning. In addition to the aforementioned directions for future research, it is important for replication studies to consider including all four tones in order to create a more authentic representation of tone use in Mandarin. A longitudinal approach would also provide more robust insights into the developmental path of Mandarin tone acquisition and enable a better understanding of the relationship between language use and the various components that affect language development.

Chapter Three

This chapter provides an overview of the study conducted within this thesis, highlighting its key aspects and outcomes. Furthermore, we explore the significance of these findings in the context of language learning, shedding light on their implications for L2 pedagogy. This chapter ends with a discussion of future research directions for the instructional application of TTS-S.

Summary of Goals and Findings

This study explored the pedagogical use of Text-to-speech-based shadowing (TTS-S) to examine its effectiveness in the L2 learning of two Mandarin tones (#1 and #4) in terms of phonological awareness, aural perception, and oral production. The research spanned a period of six weeks within a self-directed learning environment. According to the study's findings, it is inconclusive to say that the self-directed use of TTS-S can effectively lead to the improvements of these three levels of phonological development. However, the results indicate that although certain aspects of phonological development, specifically those assessed through the ABX2 and production tests, demonstrated individual progress among the participants, the overall impact did not achieve statistical significance.

Implications for L2 Pedagogy

As highlighted in the previous chapter, teachers often lack the time necessary to give their students individualized and pertinent L2 input and output opportunities (Timperley & Robinson, 2000). In response to this limitation, the proposed tool, TTS-S, offers a promising solution by employing user-friendly techniques that can be easily incorporated within the classroom environment. By leveraging TTS-S, L2 students have the potential to enhance their sound awareness, aural perception, and oral production skills in the target L2 and consequently improve their overall pronunciation abilities. According to Celce-Murcia's framework (2010), it

is crucial to provide L2 learners with opportunities to develop these three levels of phonological development in order to enhance their pronunciation abilities.

When it comes to improving L2 pronunciation, the combination of shadowing techniques (Foote & McDonough, 2017; Zajdler, 2020) with text synthesizers offers several advantages. TTS enables students to conveniently interact with computers anytime, anywhere at their own pace (Cardoso, 2018; Cardoso, 2022). This flexibility allows learners to receive exposure to accurate models of pronunciation (via TTS), and orally practice these models using traditional and pedagogically popular listen-and-repeat techniques (Shadowing). In essence, the combination of TTS and shadowing allows learners to enhance their pronunciation skills through exposure to input and ample practice opportunities.

Future Research Directions

To validate the pedagogical efficacy of TTS-S and its applicability for language learning, additional research is required. To gain a comprehensive understanding of how Mandarin tones are actually utilized in the L2, it is crucial for replication studies to include all four tones. This move would enable a more accurate depiction of the intricacies involved in the process of acquiring Mandarin tones, and it would allow for a more comprehensive analysis of learners' abilities to differentiate and produce each tone accurately, shedding light on potential challenges and areas that require further attention in L2 Mandarin tone acquisition.

Another important direction for future research is the adoption of a longitudinal approach for data collection, which could provide valuable insights into the developmental trajectory of Mandarin tone acquisition, allowing for a comprehensive examination of how learners progress over time. Through a longitudinal study, researchers can explore how Mandarin learners' perception and production of tones evolve over an extended period, identify potential challenges

or facilitative factors, and gain a deeper understanding of the developmental processes involved. In this context, researchers could also investigate how L2 learners interact with the target TTS-S technology and examine the impact of its pedagogical use.

Finally, conducting this study on a larger scale by recruiting a larger number of participants could greatly enhance its validity and generalizability. As Porte and McManus (2019) emphasize, replication is essential in confirming and reinforcing the reliability and validity of linguistic research, ultimately improving research procedures and outcomes.

Conclusion

As the world evolves, the demand for new CALL pedagogical tools becomes increasingly apparent. This study aimed to explore the potential of a relatively recent technology called TTS-S, which combines a cutting-edge tool (TTS) with a traditional classroom technique (Shadowing) to facilitate the learning of L2 pronunciation. The study focused on enhancing sound awareness, aural perception, and oral production of tones in Mandarin Chinese.

The results indicated that while there were some improvements in the overall pronunciation of tones by some participants, particularly in relation to two of the tests (namely ABX2 and production), these advancements did not reach statistical significance. Despite the promising nature of TTS-S as an innovative approach, further research is required to fully understand its effectiveness and potential impact on L2 pronunciation development.

References

Baese-Berk, M. M. (2019). Interactions between speech phonological perception and production during learning of novel phonemic categories. *Attention, Perception, & Psychophysics*, *81*(4), 981–1005. https://doi.org/10.3758/s13414-019-01725-4

Bione, T., & Cardoso, W. (2020). Synthetic voices in the foreign language context. *Language Learning & Technology, 24*(1), 169-186. https://doi.org/10125/44715

Cardoso, W. (2018). Learning L2 pronunciation with a text-to-speech synthesizer. In P. Taalas., J. Jalkanen., L. Bradley., & S. Thouësny (Eds.), *Language Learning as Exploration and Encounters; Selected papers from EuroCALL* (pp. 16-21). Research-publishing.net. https://doi.org/10.14705/rpnet.2018.26.806

Cardoso, W. (2011). The development of Coda Perception in Second language phonology: A variationist perspective. *Second Language Research, 27*(4), 433–465. https://doi.org/10.1177/0267658311413540

Cardoso, W., Collins, L., & Cardoso, W. (2021) Developmental Sequences in Second Language Phonology: Effects of Instruction on the Acquisition of Foreign sC Onsets. *Frontiers in Communication, 6*, 662934. https://doi.org/10.3389/fcomm.2021.662934

Cardoso, W. (2022). Technology for speaking development. In T.M. Derwing., M.J. Murray., & R.I. Thomson (Eds.), *Routines in L2 Pragmatics Research* (pp. 299–313). Routledge. https://doi.org/10.4324/9781003022497-21

Celce-Murcia, M., Brinton, D., & Goodwin, J. M. (2010). *Teaching pronunciation*. Cambridge University Press.

Chen, X., Ku, Y. M., Koyama, E., Anderson, R. C., & Li, W. (2008). Development of Phonological Awareness in bilingual Chinese children. *Journal of Psycholinguistic Research*, *37*(6), 405–418. https://doi.org/10.1007/s10936-008-9085-z

Collins, L., & Muñoz, C. (2016). The foreign language classroom: Current perspectives and future considerations. *The Modern Language Journal*, *100*(1), 133–147. https://doi.org/10.1111/modl.12305

Craig, S.D., & Schroeder, N. L. (2017). Reconsidering the voice effect when learning from a virtual human. *Computers & Education*, *114*, 193-205. https://doi.org/10.1016/j.compedu.2017.07.003

DeKeyser, R. M. (2013). Age effects in second language learning: Stepping stones toward better understanding. *Language Learning*, *63*(s1), 52-67. https://doi.org/10.1111/j.1467-9922.2012.00737.x

Du, F. (2013). Student Perspectives of Self-Directed Language Learning: Implications for Teaching and Research. *International Journal for Scholarship of Teaching and Learning*, *7*(2), 1–18. https://doi.org/10.20429/ijsotl.2013.070224

Ellis, N., & Collins, L. (2009). Input and second language acquisition: The roles of frequency, form, and function introduction to the special issue. *The Modern Language Journal, 93* (3), 329–335. https://doi.org/10.1111/j.1540-4781.2009.00893.x

Fang, C. (2017). *A Prototype Text Analyzer for Mandarin Chinese TTS sSystem* (Publication No. 1111826.) [Masters' thesis, Uppsala University]. DiVa Portal. https://www.diva-portal.org/ smash/get/diva2:1111826/FULLTEXT01

Foote, J. A., & McDonough, K. (2017). Using shadowing with mobile technology to improve L2 pronunciation. *Journal of Second Language Pronunciation*, *3*(1), 34–56. https://doi.org/10.1075/jslp.3.1.02foo

Gilakjani, A. P., Leong, L.M., & Sabouri, N. B. (2012). A Study on the Role of Motivation in Foreign Language Learning and Teaching. I.J. *Modern Education and Computer Science, 7*, 9-16. https://doi.org/10.5815/ijmecs.2012.07.02

Guo, L., & Tao, L. (Eds.). (2008). *Tone Production in Mandarin Chinese by American Students: A Case Study*. 20th North American Conference on Chinese Linguistics (NACCL-20).

Hamada, Y. (2018). Shadowing: What is it? how to use it. Where will it go? *RELC Journal*, *50*(3), 386–393. https://doi.org/10.1177/0033688218771380

Hao, Y. C. (2012). Second language acquisition of mandarin chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics, 40*(2), 269–279. https://doi.org/10.1016/j.wocn.2011.11.001

Hautus, M. J., & Meng, X. (2002). Decision strategies in the ABX (matching-to-sample) Psychophysical Task. *Perception & Psychophysics*, *64*(1), 89–106. https://doi.org/10.3758/bf03194559

Hendry, C. (2023). *Using Translation Tools for L2-Learning in a Self-Regulated Environment* (Publication No 991740.) [Doctoral dissertation, Concordia University]. https://spectrum.library.concordia.ca/id/eprint/991740/

Huang, Y., & Lawless, H. (1998). Sensitivity of the ABX discrimination test. *Journal of Sensory* Studies, 13(2), 229–239. https://doi.org/10.1111/j.1745-459x.1998.tb00085.x

Jurafsky, D., & Martin, J. H. (2000). *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition*. Prentice Hall.

Kachlicka, M., Saito, K., & Tierney, A. (2019). Successful second language learning is tied to robust domain-general auditory processing and stable neural representation of sound. *Brain and Language, 192*, 15–24. https://doi.org/10.1016/j.bandl.2019.02.004

Koedel, C., & Betts, J. (2010). Value Added to What? How a Ceiling in the Testing Instrument Influences Value-Added Estimation. *Education Finance and Policy, 5*(1), 54-81. https://doi.org/10.1162/edfp.2009.5.1.5104

Krashen, S.D. (1982). *Principles and practice in second language acquisition*. Pergamon Press.

Lambert, Sylvie. (1994). *Bridging the Gap: Empirical Research in Simultaneous Interpretation*. John Benjamins Publishing.

Li, Y., Tang, C., Lu, J., Wu, J., & Chang, E. F. (2021). Human cortical encoding of pitch in tonal and non-tonal languages. *Nature Communications, 12*(1), 1-12. https://doi.org/10.1038/s41467-021-21430-x

Liakin, D., Cardoso, W., & Liakina, N. (2015). Learning L2 pronunciation with a mobile speech recognizer: French /y/. *CALICO Journal, 32*(1), 1-25. https://doi.org/10.1558/cj.v32i1.25962

Liakin, D., Cardoso, W., & Liakina, N. (2017). Mobilizing Instruction in a Second-Language Context: Learners' Perceptions of Two Speech Technologies. *Languages, 2*(3), 11. https://doi.org/10.3390/languages2030011

Lin, W. (1985). Teaching Mandarin tones to adult English speakers: Analysis of difficulties with suggested remedies. *RELC Journal, 16*(2), 31–47. https://doi.org/10.1177/003368828501600207

Little, D. (1995). Learning as dialogue: The dependence of learner autonomy on teacher autonomy. *System, 23*(2), 175–181. https://doi.org/10.1016/0346-251X(95)00006-6

Liu, H. (2015). Learner Autonomy: The Role of Motivation in Foreign Language Learning. *Journal of Language Teaching and Research, 6*(6), 1165-1174. http://dx.doi.org/10.17507/jltr.0606.02

Ma, Z., & Ma, R. (2012). Motivating Chinese Students by Fostering Learner Autonomy in Language Learning. *Theory and Practice in Language Studies, 2*(4), 838-842. http://dx.doi.org/10.4304/tpls.2.4.838-842

Moon, D. (Eds.). (2012). *Web-Based Text-to-Speech Technologies in Foreign Language Learning: Opportunities and Challenges*. EL 2012 DTA. https://doi.org/10.1007/978-3-642-35603-2_19

Mueller, J. L., Friederici, A. D., & Männel, C. (2012). Auditory phonological perception at the root of language learning. *Proceedings of the National Academy of Sciences, 109*(39), 15953–15958. https://doi.org/10.1073/pnas.1204319109

Orton, J., & Scrimgeour, A. (2019). *Teaching Chinese as a Second Language: The Way of the Learner*. Routledge. https://doi.org/10.4324/9781351206877

Pennington, M. C. (2021). Teaching pronunciation: The state of the art. *RELC Journal, 52*(1), 3–21. https://doi.org/10.1177/00336882211002283

Pienemann, M. (1998). *Language processing and second language development: Processability Theory*. John Benjamins.

Porte, G. K., & McManus, K. (2019). *Doing replication research in applied linguistics*. Routledge.

Robinson, P. (2010). Individual Differences and Instructed Language Learning. *Language Learning, 60*(s2), 46-84. https://doi.org/10.1075/lllt.2

Schmidt, R. (2001). Attention. In P. Robinson (Eds.), *Cognition and second language instruction,* (pp.3-32). Cambridge University Press.

Shen, G., & Froud, K. (2018). Electrophysiological correlates of categorical perception of lexical tones by English learners of Mandarin Chinese: An ERP study. *Bilingualism: Language and Cognition, 22*(2), 1-13. http://dx.doi.org/10.1017/S136672891800038X

Shao, Y., Saito, K., & Tierney, A. (2022). How does having a good ear promote instructed Second language pronunciation development? roles of domain-general auditory processing in choral repetition training. *TESOL Quarterly, 43*(3),581-606. https://doi.org/10.1002/tesq.3120

Shiki, O., Mori, Y., Kadota, S., & Yoshida, S. (2010). Exploring differences between shadowing and repeating practices: An analysis of reproduction rate and types of reproduced words. *Annual Review of English Language Education in Japan*, *21*, 81–90. https://doi.org/10.20581/arele.21.0_81

Swain, M. (2005). The output hypothesis: theory and research. In E. Hinkel (Eds.), *Handbook of research in second language teaching and learning* (pp. 471-483). Lawrence Erlbaum. https://doi.org/10.4324/9781410612700

Timperley, H., & Robinson, V. (2000). Workload and the professional culture of teachers. *Educational Management & Administration*, *28*(1), 47–62. https://doi.org/10.1177/0263211x000281005

Thornton, K. (2010). Supporting Self-Directed Learning: A Framework for Teachers. Language *Education in Asia, 1*(1), 158-170. http://dx.doi.org/10.5746/LEiA/10/V1/A14/Thornton

Tavakoli, H. (2013). *A dictionary of research methodology and statistics in applied linguistics*. Rahnamā.

Triskova, H. (2017). Acquiring and teaching Chinese pronunciation. In I. Kecskes (Eds.), *Explorations into Chinese as a second language* (pp. 3–30). Springer. http://dx.doi.org/10.1007/978-3-319-54027-6_1

Van Lieshout, C., & Cardoso, W. (2022). Google Translate as a Tool for Self-Directed Language Learning. *Language Learning and Technology, 26*(1), 1-19. http://hdl.handle.net/10125/73460

Wang, M., Perfetti, C., & Liu, Y. (2003). Alphabetic readers quickly acquire orthographic structures in learning to read Chinese. *Scientific Studies of Reading, 7*(2), 183–208. https://doi.org/10.1207/S1532799XSSR0702_4

Wu, X., & Lin, H. (2008). Perception of Mandarin Tones by Mandarin and English Listeners. *Journal of Chinese Language and Computing, 18*(4), 175-187.

Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics, 33*(2), 159-197. https://doi.org/10.1016/j.wocn.2004.11.001

Yang, B. (2016). *Perception and production of Mandarin tones by native speakers and L2 learners*. Springer Berlin.

Yang, C. (2021). Categorical Perception of Mandarin Tones by Native and Second Language Speakers. In X. L. Zhang & Y. Chen (Eds.), *The Acquisition of Chinese as a Second Language Pronunciation* (pp. 55-74). Springer. http://dx.doi.org/10.1007/978-981-15-3809-4_3

Yang, J., & Medwell, J. (2019). Learners' and Teachers' Beliefs About Learning Tones and Pinyin. In I. Kecskés (Eds.), *Explorations into Chinese as a second language* (pp. 141–163). SPRINGER. http://dx.doi.org/10.1007/978-3-319-54027-6_6

Yanguas, I. (2012). Task-based oral computer-mediated communication and L2 vocabulary acquisition. *CALICO, 29*(3), 507–531. https://doi.org/10.11139/cj.29.3.507-531

Zajdler, E. (2020). Speech shadowing as a teaching technique in the CFL classroom. *Lingua Posnaniensis, 62*(1), 77–88. https://doi.org/10.2478/linpo-2020-0005

Zhang, Y., & Gao, X. (2022). From individual tones to connected speech: The case of tone 3. In Y. Zhang & X. Gao (Eds.), *Frontiers of L2 Chinese language education a global perspective* (pp. 30–50). Routledge.

Appendix A

Sample of the texts presented to the participants

Week 1 - Text 1

1. 你是中国人吗？
2. 我是加拿大人。你呢？
3. 他是北京人。
4. 她做什么工作？
5. 高先生是老师。
6. 这个人是谁?
7. 这是我妈妈。
8. 这是我爸爸。

1. Ni shi zhongguo ren ma.
2. Wo shi jianada ren. Ni ne?
3. Ta shi beijing ren.
4. Ta zuo shenme gongzuo?
5. Gao xiansheng shi laoshi.
6. Zhe shi shei?
7. Zhe shi wo mama.
8. Zhe shi wo baba.

Translation:

1. Are you Chinese?
2. I am Canadian. You?
3. He is from Beijing.
4. What is her job?
5. Mr. Gao is not a teacher.
6. Who is this person?
7. This is my mother.
8. This is my dad.

Appendix B

Demographic Questionnaire

Participant's code : _____

1. Age: 18-20 21-25 26-30 31-35 36-40 41 and up
2. Gender (optional): _____
3. What is/are your first language/s?

   _____

4. What other languages do you know or speak?

   _____

5. Have you ever learned any languages on your own? Yes / No
    1. If so, which one(s)? _____
    2. If so, what tools/books/websites did you use to learn?

       _____

    3. If so, what tools/books/websites did you use to learn pronunciation? Which?

       _____

6. Have you ever used shadowing? — Shadowing consists of the repetition of a word, phrase, or sentence immediately after hearing it.
7. If you have used shadowing, what do you use it for?

   _____

8. Have you ever used text-to-speech software (text readers)?
    1. If you have used text-to-speech software, what do you use it for?

       _____

    2. Have you ever used 'Ispeech'? Yes No

9. What is your highest level of education?

Primary Secondary Bachelor’s Master's PhD

What do/did you study?

_____

10. I enjoy using technology to learn languages.

1 2 3 4 5 6

(Not much) (Very much)

11. How interested are you in learning new languages?

1 2 3 4 5 6

(Not much) (Very much)

12. I am motivated to learn Mandarin in this study.

1 2 3 4 5 6

(Strongly Disagree) (Strongly Agree)

13. Have you ever learned or been exposed to the Mandarin language? If so, please explain.

_____

_____

_____

_____

Appendix C

Awareness test

1. ma1 ma4
    1. Are those two sounds the same?
        1. Yes 2. No
    2. Explain.
2. Mi4, mi1
    1. Are those two sounds the same?
        1. Yes 2. No
    2. Explain.
3. me1 me1
    1. Are those two sounds the same?
        1. Yes 2. No
    2. Explain.
4. Mu1 mu4
    1. Are those two sounds the same?
        1. Yes 2. No
    2. Explain.
5. mo4, mo4
    1. Are those two sounds the same?
        1. Yes 2. No
    2. Explain.
6. mi1 mi1

1. Are those two sounds the same?

    1. Yes 2. No

2. Explain.

Appendix D

ABX 1 Pretest and Posttest

1. ma1 VS ma4
2. mo4 VS mo4
3. mu4 VS mu1
4. ma1 VS ma4
5. me1 VS me1
6. mi4 VS mi1
7. mi1 VS mi4
8. mo1 VS mo1
9. mu4 VS mu1
10. me4 VS me4

Appendix E

ABX 2 Pretest and posttest

1. Control sound: ma1 VS ma4, Blind sound: ma1
2. Control sound: me1, me4, Blind sound: me4
3. Control sound: mi1, mi4, Blind sound: mi4
4. Control sound: mi1, mi4, Blind sound: mi1
5. Control sound: mi1, mi4, Blind sound: mi4
6. Control sound: mo1, mo4, Blind sound: mo4
7. Control sound: mo1, mo4, Blind sound: mo1
8. Control sound: mo1, mo4, Blind sound: mo4
9. Control sound: ma1 ma4, Blind sound: ma4
10. Control sound: me1, me4, Blind sound: me1
11. Control sound: me1, me4, Blind sound: me1
12. Control sound: mu1, mu4, Blind sound: mu1
13. Control sound: ma1 ma4, Blind sound: ma4
14. Control sound: mu1, mu4, Blind sound: mu4
15. Control sound: mu1, mu4, Blind sound: mu1
16. Control sound: ma1 ma4, Blind sound: ma1
17. Control sound: me1, me4, Blind sound: me4
18. Control sound: mi1, mi4, Blind sound: mi1
19. Control sound: mo1, mo4, Blind sound: mo1
20. Control sound: mu1, mu4, Blind sound: mu4

Appendix F

Example of the answer sheet provided to the test-takers

Participant number: _____

Section 1: Are the two sounds the same or are they different?

1. Same Different

If the sounds are different, what is the difference between the two sounds?

_____

_____

2. Same Different

If the sounds are different, what is the difference between the two sounds?

_____

_____

3. Same Different

If the sounds are different, what is the difference between the two sounds?

_____

_____

4. Same Different

If the sounds are different, what is the difference between the two sounds?

_____

_____

5. Same Different

If the sounds are different, what is the difference between the two sounds?

_____

_____

6. Same Different

If the sounds are different, what is the difference between the two sounds?

_____

_____

Section 2: Are the two sounds the same or are they different?

1. Same Different
2. Same Different
3. Same Different
4. Same Different
5. Same Different
6. Same Different
7. Same Different
8. Same Different
9. Same Different
10. Same Different

Section 3: You will be presented with two sounds (A and B) followed by a third one (Sound C). Identify the Sound C as either Sound A or Sound B.

1. Sound A Sound B
2. Sound A Sound B
3. Sound A Sound B
4. Sound A Sound B
5. Sound A Sound B
6. Sound A Sound B
7. Sound A Sound B
8. Sound A Sound B
9. Sound A Sound B
10. Sound A Sound B
11. Sound A Sound B
12. Sound A Sound B
13. Sound A Sound B
14. Sound A Sound B
15. Sound A Sound B
16. Sound A Sound B
17. Sound A Sound B
18. Sound A Sound B
19. Sound A Sound B
20. Sound A Sound B

Appendix G

Example pretest and posttest Production task

The participants were instructed to orally repeat (listen-and-repeat) the sound they heard after it was produced:

List of pseudowords (syllables) to produce:

1. Ma1
2. Me4
3. Mi4
4. Mo4
5. Me1
6. Ma4
7. Mi1
8. Mu4
9. Mu1
10. Mo1