Beyond the walls of classrooms: Exploring the pedagogical effectiveness of Text-To-Speech-based Shadowing (TTS-S) on the development of Mandarin tones

Sue-Anne Richer[a] and Walcir Cardoso[b]

[a]Department of Education, Concordia University, Montréal, Canada, [ID], sue-anne.richer@mail.concordia.ca and [b]Department of Education, Concordia University, [ID], walcir.cardoso@concordia.ca

Abstract

This study examines the pedagogical effectiveness of using Text-To-Speech synthesis (TTS) combined with Shadowing (TTS-S) for self-regulated learning of Mandarin tones 1 and 4. The aim is to determine the probability of success of this innovative approach that uses TTS to generate audio for shadowing practice. The research was guided by the following research question: can TTS-S help L2 learners raise their awareness and improve their perception and production of the target Mandarin tones over a period of six weeks? Over six week, ten participants engaged in self-regulated activities using TTS-S to learn the two pronunciation targets. By means of pre-/post-tests (to assess effectiveness in pronunciation) participants were asked to complete: (1) an awareness task in which they verbalized their metacognitive knowledge of Mandarin tones; (2) ABX tasks to assess their perception of Mandarin tones; and (3) a production task to evaluate their production of the target tones. Our findings are inconclusive regarding the effectiveness of TTS-S for improving awareness, perception, and production of Mandarin tones 1 and 4 among L2 learners. They also indicate that while certain aspects of phonological development, specifically production, showed some improvements, the overall impact was not statistically significant.

Keywords: Text-To-Speech synthesis (TTS), shadowing, L2 pronunciation, speech perception.

1. Introduction

With limited classroom time (Collins & Muñoz, 2016), teachers struggle to provide personalized language input (e.g. listening activities) and opportunities for students to practice oral 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' aural skills and provide flexible language practice (Little, 1995). Although TTS offers many benefits (e.g. immediate access to the language; Liakin et al., 2017), it lacks an output-inducing component that promotes active speech production (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 (e.g. 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 (whereby learners set goals for their learning and then attempt to monitor and regulate their progross) 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 test to evaluate the participants' metacognitive knowledge of Mandarin tones; (2) two ABX tests to assess their perception of Mandarin tones; and (3) a production test to evaluate their ability to produce the target tones.

2. Method

2.1. Participants and procedures

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. 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. 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 two of these tones are. The participants were then provided with five texts containing five to nine 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). On a weekly basis (Weeks 1-5), the participants learned the target Mandarin tones by: (1) copying and pasting these texts into Ispeech (ispeech.org; the TTS application adopted); (2) listening to the synthesized output; and (3) immediately repeating it using the proposed shadowing technique (TTS-S). If they decided to use additional material, they were asked to add it to their weekly learning logs. 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. The posttests were administered one week after the completion of the self-directed treatment proposed.

2.2. Instruments and data analysis

To evaluate the phonological development of tones, this study looked at the effects of the treatment on the sound awareness, aural perception, and oral production of tones in a self-directed TTS-S based environment. A sound awareness test was used to measure the metacognitive knowledge of the students about tones in Mandarin. Six minimal pairs (e.g. ma1 vs ma4) were played and the participants were asked to label the two sounds as the same or different and then, if applicable, explain how the sounds differ.

The aural perception tests used two ABX aural discrimination tests (for the rationale, see Hautus & Meng, 2002). In the first ABX tests, participants were asked to determine whether there was an audible difference between two audio signals. In the second ABX test, participants followed a match-to-sample discrimination procedure: two sounds (test and control) were presentend to the participants, followed by a third sound (blind sample) that corresponded to the test sound. The participants were asked to match the blind sample to the test item.

Because of the low proficiency of the participants and the short duration of the treatment, the production test adopted a 'listen and repeat' activity (for the rationale, see Cardoso et al, 2021), a less demanding test in comparison with reading aloud and spontaneous conversation. This test asked the participants to listen to a recording of a native speaker pronouncing a word in Mandarin and to repeat it as accurately as possible.

The phonological awareness data were analyzed via a 3-level system: 0 (fully incorrect = not aware); 1 (partially

correct = partial knowledge); and 2 (fully correct = full knowledge). The perception and production data were analyzed using descriptive statistics (e.g. means-M, standard devision-SD) via SPSS. All scores were converted to a 100% scale for comparison purposes. Means of pretests and posttests were compared using paired sample t-tests.

3. Results

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 the target Mandarin tones in terms of sound awareness, aural perception, and oral production. The paired sample t-test results indicate that the posttest means for all tests were higher compared to the means obtained in pretests. However, these differences were not statistically significant. Table 1 provides a summary of the descriptive 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 Gains	SD
Awareness	Pretest Posttest	88.34 98.34	10.43 2.64	5	10.54
ABX1	Pretest Posttest	97 99	4.83 3.16	2	4.21
ABX2	Pretest Posttest	88 94.5	16.36 8.31	6.5	11.56
Production	Pretest Posttest	79.75 82	15.47 15.47	2.25	5.83

An interesting pattern observed across all tests is that some participants demonstrated improved knowledge of Mandarin tones (e.g. Participants 2, 6, and 10 in Figure 1), while others showed no change. Due to space limitation, we only provide the results for oral production, shown in Figure 1.
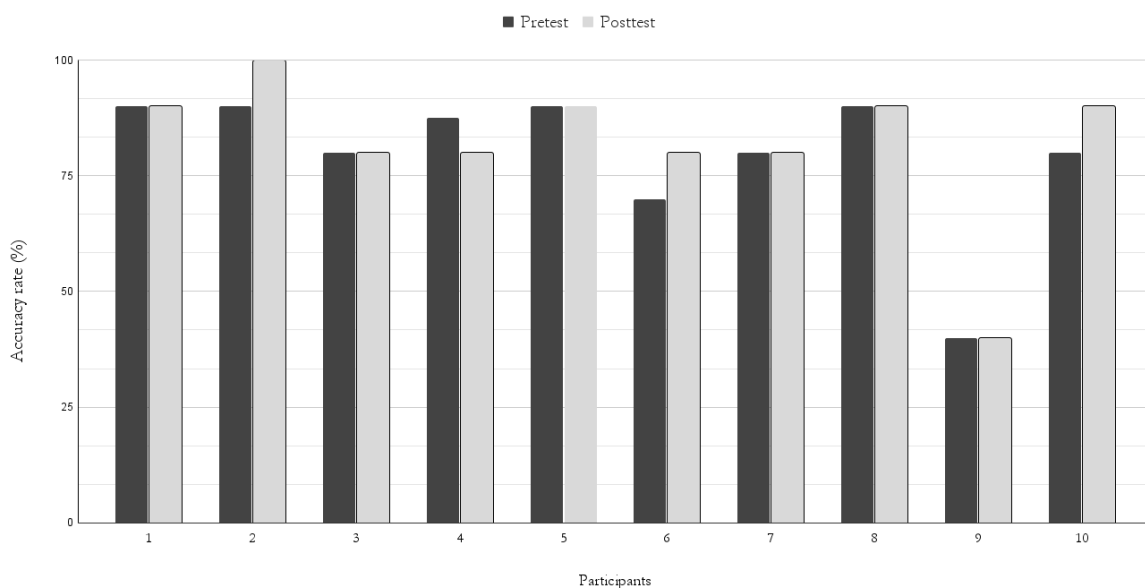

Figure 1. Production Task: Accuracy rate (%) by participant.

4. Discussion

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, this study 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 the improvements observed in all four tests (i.e. sound awareness, aural perception, and oral production) were not statistically significant. This suggests that, overall, the proposed treatment did not yield any significant improvements in the assessed measures of pronunciation development.

Two important generalizations 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). For instance, in general we observed that participants exhibited a high level of awareness and perception regarding tone 1 and tone 4, suggesting a 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 (Robinson, 2010). The varying degrees of improvement among participants highlights how personal factors can influence the effectiveness of our proposed intervention, with some benefiting more than others.

5. Conclusions

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. Importantly, the small sample size and the individual differences among the participants should be taken into consideration when interpreting these findings. 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.

It is worth acknowledging some of the limitations of this study, which could also have influenced the results. A key limitation is its small sample size of only ten 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 paper), larger samples and individual interviees 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, the listen-and-repeat oral production test), 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.

Acknowledgements

References

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.), *The Routledge Handbook of Second Language Acquisition and Speaking* (pp. 299–313). Routledge. https://doi.org/10.4324/9781003022497

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

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

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

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

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

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

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

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.

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