Mandarin speakers prefer explicit visual cues in learning Cantonese tones: an eye-tracking study

Yuqin Shu, Yi Weng, Ran Tao, Gang Peng

Proceedings of the 38th Pacific Asia Conference on
Language, Information and Computation (PACLIC 38)

Nathaniel Oco, Shirley N. Dita, Ariane Macalinga Borlongan, Jong-Bok Kim (eds.)

2024

Mandarin speakers prefer explicit visual cues in learning Cantonese tones: an eye-tracking study

Yuqin Shu, Yi Weng, Ran Tao, Gang Peng
Research Centre for Language Cognition, and Neuroscience, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China
yuqin.shu@connect.polyu.hk, yvonne-yi.weng@connect.polyu.hk, ran.tao@polyu.edu.hk, gang.peng@polyu.edu.hk

Abstract

This study investigates the role of visual feedback on Mandarin speakers learning Cantonese tones using a high-variability perceptual learning paradigm. Thirty Mandarin speakers participated in a two-day experiment, completing pre-tests, training, post-tests, and generalization tests. Explicit (tone letters) and implicit (tone numbers) information related to tones were provided during training. Participants' eye movements were recorded during training. The testing results showed that the identification of Cantonese tones by Mandarin speakers improved significantly, demonstrating the effectiveness of the training procedure incorporating visual feedback. Eye-tracking data revealed that participants spent the most time fixating on tone letters, and their attention to these letters increased as the training progressed. These findings highlight the importance of explicit visual information in auditory perceptual learning of tones. The impact of Mandarin tone experience on learning Cantonese tones was also discussed.

1 Introduction

Different languages utilize acoustic cues differently. In tonal languages, lexical tones serve to differentiate the meanings of words. It is thus essential for learners to learn to correctly identify the tones from different categories to understand the meaning of words. However, mastering novel lexical tone categories has been found challenging not only for nontonal speakers who have little tone experience but also for tonal speakers whose acoustic features of the native tones differ from that of the novel tonal language (So & Best, 2010; Francis et al., 2008; Hao, 2012). Taking Mandarin and Cantonese for example, there are four lexical tones in Mandarin (high level tone T55, high rising tone T35, low falling-rising tone T214, and high falling tone T51) but six tones in Cantonese (high level tone T55, high rising tone T25, middle level T33, low falling tone T21, low rising tone T23, and low level toneT22). Peng (2006) displayed the distributions of Mandarin and Cantonese tones in a two-dimension coordinate where the x-axis was pitch slope while the y-axis was pitch height, showing that the tone balloons for Mandarin tones were compact and discretely distributed yet there were overlaps among balloons for Cantonese tones. Such acoustic features have led to difficulties for Mandarin speakers to perceive Cantonese tones (Zhang et al., 2016).

The effectiveness of perceptual learning paradigm in tonal learning has been widely recognized (Wang et al., 1999; Chandrasekaran et al., 2010; Francis et al., 2008). In recent years there has been an increasing number of studies focusing on the role of multimodal information in auditory perceptual learning, in which visual information plays an important role. According to the dual-coding theory (Paovoi, 1986) and the cognitive theory of multimedia learning (Mayer, 2001), the mental representation of speech sounds would become more robust if information is presented through both auditory and visual channels, contributing to better learning outcomes. In terms of tone learning, many previous studies have found that various kinds of visual information can facilitate learner's ability to perceive the tones correctly, such as real pitch contours of the tone (Liu et al., 2011), static or dynamic pitch changes (Wei et al., 2022; Godfroid et al., 2017; Zhen et al., 2019), hand gestures (Morett et al., 2022; Morett &

Chang, 2015; Baills et al., 2019), numbers (Liu et al., 2011; Godfroid et al., 2017) and colors (Godfroid et al., 2017). The visual facilitation mentioned above can be divided into two kinds, one is the explicit information that gives a direct indication of the pitch height or direction of the tone such as pitch contour, arrows, etc.; and the other is implicit information that does not cue the pitch-acoustic change of the tone but only provides a way to label it, such as a number or a color. Although a few studies have compared the learning effects under different visual information aids (Godfroid et al., 2017), it is still unclear which type of information, namely, explicit or implicit, learners would prefer when both are provided for them to opt for.

Another frequently studied but unresolved issue is the potential impact of native tone experience on learning new tonal languages. Previous studies on Mandarin learners' perception of Cantonese tones have reached some consensus: after training, T55 and T21 in Cantonese are better distinguished, while the two level tones (T22 and T33) are very difficult to identify (Francis et al., 2008; Zhang et al., 2008; Chang et al., 2017; Jongman et al., 2017). However, there is disagreement regarding the specific tone confusion patterns. Zhang et al. (2016) found that Mandarin speakers' tone confusion primarily occurs bidirectionally between tones sharing a similar pitch contour, such as T22-T33 (the two level tones) and T23-T25 (both rising tones), with little confusion among other tones. In contrast, Francis et al. (2008) found additional confusions mainly induced by pitch height, including misidentifying T22 as T21 more frequently than as T33 and confusing T23 and T21 bidirectionally. These two confusion patterns represent different influences of native language experience. The former suggests that confusion occurs only along the pitch height dimension, indicating that pitch contour might be a more dominant cue to suppress confusion. In contrast, the latter suggests that confusion exists along both the pitch height and pitch contour dimensions.

Learning Cantonese tones by Mandarin speakers provides an opportunity to explore both visual preferences and the influence of native language on nonnative tone acquisition. Mandarin speakers are well experienced in learning Mandarin tones with the aids of both explicit symbols and implicit numbers. Starting as early as the first grade of elementary school, Mandarin speakers have been systematically taught Pinyin, the phonetic symbols for Chinese characters, in which tones are named by 1 to 4 and are depicted by contour markers above the vowels. For instance, “mā”, “má”, “mǎ”, “mà” indicate syllable “ma” with tone 1 to 4. Such extensive experience in using both explicit and implicit cues may lead Mandarin speakers to have a balanced preference for explicit and implicit visual information when learning Cantonese tones. Additionally, examining the confusion patterns in Cantonese tone perception by Mandarin speakers adds insights into how native tone language speakers learn nonnative tones and the influence of their native language on this process.

In this study, we adopt the high-variability perceptual learning paradigm that provides visual feedback to train Mandarin speakers to learn Cantonese tones. Specifically, we provide both explicit and implicit visual information related to tones, and we are mainly concerned with the following two questions: 1) whether Mandarin speakers prefer implicit or explicit information when they acquire new tones in another language (i.e., Cantonese), and 2) How does their native tonal language background influence their acquisition of Cantonese tones?

2 Methodology

Participants

30 native Mandarin speakers (17 female, mean age = 24.3 yrs, SD = 2.13) were recruited to participate in the experiment. All of them are college students in Hong Kong, with no self-reported visual, hearing, or cognitive impairment. One male participant was left-handed. The participants resided in Hong Kong for an average period of 9.2 months (SD = 6.10) and none of them had previous knowledge of Cantonese. All participants signed written consent before the experiment. The experiment protocol was approved by the Human Subjects Ethics Sub-committee of The Hong Kong Polytechnic University.

Stimuli

Stimuli of this study were 24 Cantonese monosyllables deriving from 4 carrier syllables (/fɐn/, /fu/, /ji/, and /sɛ/) × 6 Cantonese tones (T55, T25, T33, T21, T25, and T22). All monosyllabic stimuli involved were real words in Cantonese. Four native Cantonese speakers (2 females) were recruited to pronounce each word three times in a

sound-attenuated booth, rendering three tokens for a word from one speaker. One token of each word from two speakers (one male and one female), was chosen as the standard sound across pretest, training, and posttest for each subject. In the generalization test, all three tokens per word, produced by the other two speakers, were utilized as stimuli to investigate the generalization of training effects derived from limited exposure to standard phonetic cues onto novel materials. All stimuli were normalized to 450ms in duration and 70 dB in intensity using Praat (Boersma & Weenink, 2024).

Experiment Procedure

The whole experiment lasted for two consecutive days and included two sessions of Cantonese tone training and three sessions of testing with one conducted before the training (i.e., pre-test) and two after the training (i.e., post-test and generalization test). Each session of the experiment used a block design, with blocks divided by male and female speakers, and the order of the blocks was counterbalanced. The training program adopted the perceptual learning paradigm and high variability phonetic training, with participants' eye movements being tracked throughout using an SR Research EyeLink 1000 Plus sampling at 1000 Hz. In testing phrases, tone identification task was used to evaluate participants perceptual accuracy of the target words. In day 1, participants completed the pre-test and first training. In day 2, they received the second training and attended the post- and generalization test immediately.

The training procedure started with a context where the sounds and corresponding Chinese character of the syllable /sɛ/ were presented sequentially with the six Cantonese tones. Then the formal training trials began, with syllables /fɐn/, /fu/ and /ji/ being the target stimuli. In each trial, participants were presented with a fixation cross (500ms), a monosyllabic stimulus (450ms), followed by a response screen with six options covering all Cantone six tone categories. Participants then made a response based on their perception by pressing the number keys from 1 to 6 on the keyboard. After that, a blue or red cross appeared on the screen to indicate the correctness of their choice, with blue denoting correctness and red denoting errors. Subsequently, a correct information display regarding the heard sound appeared, presenting four types of information for three seconds (i.e., the four Areas of Interests, AOIs): 1) tone number, numbers from 1 to 6 which indicate the Cantonese tone categories; 2) tone letter, consisting of a vertical bar representing the range of possible pitch heights and a branching bar representing the onset and offset of pitch heights of a tone (Chao, 1930); 3) character of the target sound and 4) English meaning of the target sound. The locations of AOIs were counterbalanced and pseudo-randomized. Before the training commenced, participants were briefed on the meanings of each type of information. Participants were instructed that they could freely choose their learning strategy. The two training sessions took about 1 hour, consisting of 288 trials (3 carrier syllables × 6 tones × 4 repetitions × 2 speakers × 2 training sessions) in total.

The procedure for the tone identification task in the three tests was very similar to the training process, with the main difference being that no feedback or information was provided. Next trial was proceeded automatically after detecting a choice. In each test, there were 108 trials, resulting from 3 carrier syllables × 6 tones × 3 repetitions × 2 speakers.

3 Results

Results for tone identification

Figure 1 illustrates the accuracy of the tone identification task in pre-test, post-test and generalization test. Accuracy results were submitted to a two-way repeated measures analysis of variance (ANOVA) with *Test* (pre-, post- and generalization test), and *Tone* as the with-in subject factors. Necessary post-hoc analyses were conducted through Tukey method for comparing families of multiple estimates. There were significant main effects of *Test* (F (2,58) = 215.2, p < 0.001). Post-hoc analysis showed that the accuracy of both post-test (72.1%) and generalization test were significantly higher than that of pretest (27.5%), ps < 0.001. Accuracy of post-test was significantly higher than that of generalization test (p = 0.04). These results showed that participants' perception of Cantonese tones was greatly improved after training and the ability to identify tones was generalized to untrained sounds to a certain degree. The main effect of *Tone* was also significant (F (5, 154) = 85.64, p < 0.001). T55 is the easiest tone to be identified with the highest overall accuracy of 79.8%. Next is T21

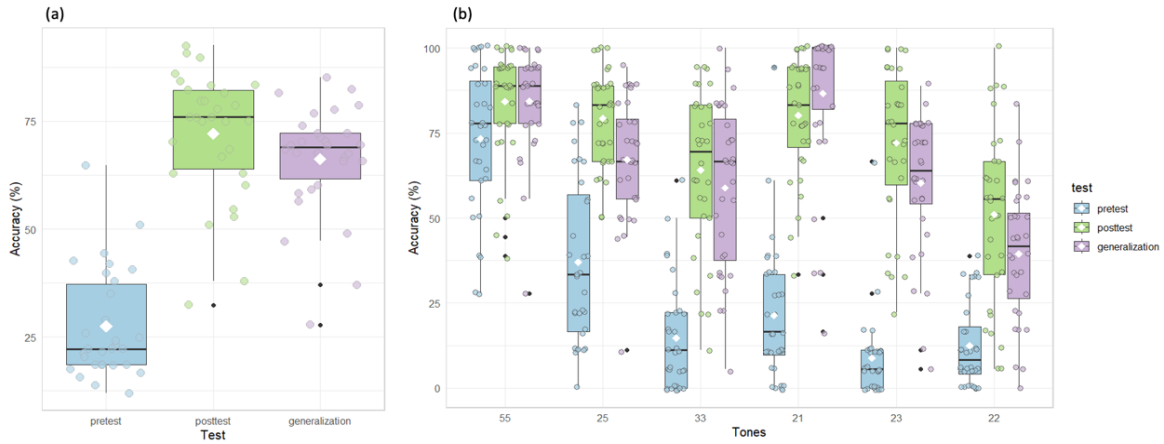

Figure 1. The accuracy of tone identification task in pre-test, post-test and generalization test. (a) shows the overall accuracy, (b) shows the accuracy of 6 Cantonese tones in three tests. The white rhombus indicates the mean value.

(a)

	R55	R25	R33	R21	R23	R22
T55	83.9%	0.9%	11.3%	0.4%	0.9%	2.6%
T25	0.0%	79.6%	0.2%	2.0%	18.1%	0.0%
T33	23.9%	0.4%	63.7%	0.7%	1.5%	9.8%
T21	0.0%	0.9%	2.6%	81.7%	3.9%	10.9%
T23	0.2%	21.7%	1.5%	2.8%	72.2%	1.7%
T22	4.3%	0.7%	39.1%	2.4%	2.0%	51.5%

(b)

	R55	R25	R33	R21	R23	R22
T55	83.3%	0.7%	11.3%	0.0%	0.9%	3.7%
T25	0.4%	67.3%	0.6%	2.2%	28.9%	0.6%
T33	21.1%	0.9%	58.9%	0.4%	2.0%	16.7%
T21	0.2%	1.9%	0.4%	88.0%	3.7%	5.9%
T23	0.6%	30.6%	2.0%	4.1%	60.6%	2.2%
T22	9.3%	0.2%	48.3%	1.1%	1.7%	39.4%

Table 1. Confusion matrix of tone identification for (a) post-test and (b) generalization test. The letter T stands for the target responses and the letter R refers to the responses given by the participants.

(63.6%) and T21 (61.2%), followed by T23 (47.0%) and T33 (45.9%), and the most difficult tone to identify is T22, with the lowest accuracy of 34.0%.

The interaction between *Test* and *Tone* (F (10, 290) = 18.58, $p < 0.001$) was also significant, suggesting that participants' ability to correctly identify tones improved differently depending on the specific tone. Specifically, apart from T55 which consistently maintained a high accuracy rate with no significant changes across the three tests, the recognition accuracy of the other five lexical tones in the post-test and generalization test was significantly higher than in the pre-test ($ps < 0.001$) with no difference between the post-test and generalization test ($ps > 0.19$). In the post-test, there were no significant differences in accuracy rates among T55, T25, T21, and T23. In comparison, the accuracy rates for T22 and T33 were significantly lower ($ps < 0.01$). By the time of the generalization test, T55 and T21 exhibited the highest accuracy rates, significantly surpassing T25, T33, and T23 ($ps < 0.01$), with T22 showing

significantly lower accuracy compared to all other tones ($ps < 0.01$).

Examination of confusion in post- and generalization tests (Table 1) provides some qualitative context for interpreting tone identification accuracy results. In both tests, T55 was also highly accurate and was only occasionally misidentified as T33 (11.3% in both tests). For T33, the mid-level tone, is consistently misidentified as T55 (23.9% and 21.1% in pre-test and generalization test) and T22 (9.8% and 16.7% respectively) across both tests. The low-level tone (T22) was the hardest one to identify across both tests. It was frequently misidentified as T33 in both tests (39.1% and 48.3% respectively) and occasionally misidentified as T55 in generalization test (9.3%). The high-rising tone (T25) and the low-rising Tone (T23) were mostly confused with each other, with a notably increasing confusion from the post-test to the generalization test (T25 misidentified as T23: from 18.1% to 28.9%; T23 misidentified as T25: from 21.7% to 30.6%). The only falling tone (T21) was maintained high accuracy with minimal confusion.

Results of fixation duration during training

To learn more about how participants allocate their attention to the four types of information during training, we analyzed the fixation duration of the participants within the 3-second time window of information display. One participant's data was identified as outlier and was excluded from the analysis. Figure 2 illustrates participants' average fixation duration. A Kruskal-Wallis rank sum test

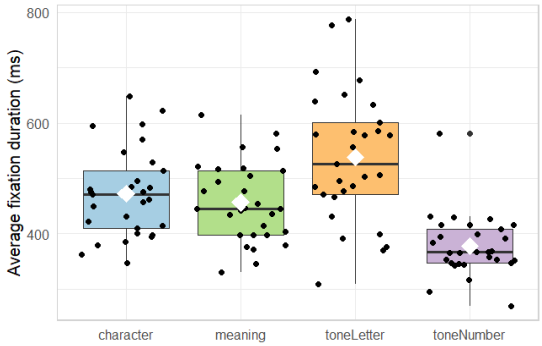

Figure 2. Average fixation duration of participants looking at the four AOIs during training.

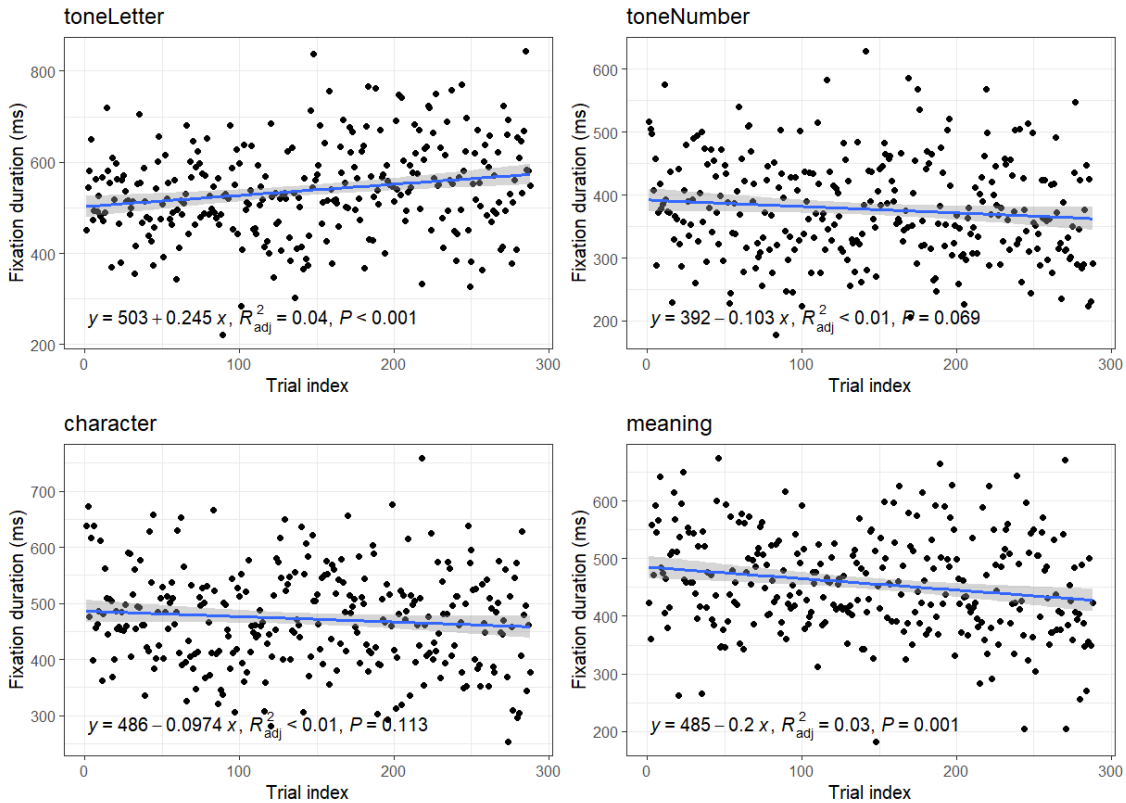

Figure 3. The changes in participants' fixation durations on the four types of information over the course of training sessions. Top left panel: Tone letter; top right panel: Tone number; bottom left panel: Chinese character; bottom right panel: English meaning.

was conducted to determine if there were statistically significant differences in average fixation duration across different AOIs. The results showed a significant difference between AOIs ($\chi^2(3) = 39.85$, $p < 0.001$), indicating that participants paid unequal attention to different information. To further investigate these differences, pairwise comparisons were performed using the Wilcoxon rank sum exact test with Bonferroni correction for multiple comparisons. The fixation duration of tone letter was significantly higher than that of meaning ($p = 0.04$) and tone number ($p < 0.001$). No significant difference was found between fixation duration of character and that of meaning ($p = 1$). Tone number got the least attention during training ($ps < 0.01$) when compared with other three types of information.

Figure 3 illustrates the fixation patterns of participants towards four AOIs during training. As can be seen, participants' attention to the tone letter significantly increased with the increasing number of trials ($\beta = 0.245$, $t(286) = 3.54$, $p < 0.001$), indicating that their reliance on the tone letter enhanced as the training progressed. Participants' attention to meaning significantly decreased along with the increase in trials ($\beta = -0.2$, $t(286) = -3.30$, $p = 0.001$), while that to character and number remained relatively stable, showing no significant changes.

4 Discussion

In this study, we trained Mandarin speakers to learn Cantonese tones, while recording their eye movements during the training process. The results showed a significant improvement in participants' perception accuracy on Cantonese tones produced by trained and new speakers, indicating the high-variability training program that provided feedback and visual indication effectively sharpened the perception of nonnative tone in learners from tonal language backgrounds. However, such learning and generalization effects were not equally manifested across all six tones, as participants, who achieved significantly higher accuracy in identifying T55 and T21, were relatively prone to mutual confusion between T25 and T23, as well as between T33 and T22. Additionally, the eye-tracking results revealed that participants had different preferences for different types of visual information, they tended to focus more on tone letters which reflected the pitch contour of tones but less on numbers that are also frequently used in Mandarin to refer to tone categories.

Mandarin speakers' preference for explicit visual information when learning nonnative tones

Contrary to our prediction that Mandarin speakers might have balanced preference for tone letters and numbers, we found that when given the freedom to choose their learning strategies and allocate their attention, Mandarin speakers spontaneously paid the most attention to the tone letters and the least attention to tone numbers. These results suggest that when Mandarin speakers were newly exposed to nonnative tones, they preferred to draw on explicit rather than implicit visual information to help reinforce the phonetic features of the tones and aid speech perception.

A possible reason for this behavior is that, since the explicit tone letters directly depict the acoustic features of tones, Mandarin learners may find it easier to guide top-down attention to enhance the integration of auditory and visual cues. In contrast, the implicit tone numbers offer limited pitch information, which might cause extra cognitive load to establish a correspondence between each number and a specific tone. Given that Cantonese has two more tones than Mandarin, this task becomes even more challenging. Besides, due to the robust correspondence between numbers (1 to 4) and Mandarin tones (T1[55], T2[35], T3[214], and T4[51]) in Mandarin speakers' memory, there might be interference with the establishment of new tonal categories through numbers, which could lead Mandarin speakers to avoid relying on numbers to learn new tones.

However, it's essential to note that due to the relatively short training duration in this study (approximately 1 hour), the observed attention patterns may only represent learners' initial exposure to a new tone system. It remains an open question whether learners will allocate more attention to other cues as training time increases.

The influence of L1 tones on the acquisition of nonnative tones

Our findings are consistent with previous studies (Francis et al., 2008; Zhang et al., 2016), showing that T55 and T21 are the easiest tones for Mandarin speakers to identify and are seldom confused with other tones. T22 is the most difficult, while T23,

T25 and T33 are intermediate in difficulty yet easily confused with other tones. Our confusion pattern aligns with that of Zhang et al. (2016), with confusion occurring mainly between T22 and T33 (the two level tones), and T23 and T25 (the two rising tones).

Mandarin speakers, who are sensitive to pitch slope, encounter greater confusion in distinguishing tones sharing similar pitch directions but varying pitch heights. As a result, even though the pitch difference between T22 and T21 is numerically smaller than that between T23 and T25, participants rarely confused them since T21 is a falling tone. This may be because changes in pitch slope are easier for Mandarin speakers to perceive and learn than pitch height, which may be influenced by the perception of tones in the native Chinese language. As described earlier, the pitch difference between the four tones of Mandarin is large, and a more notable feature is that each tone has a distinctive pitch contour, therefore Mandarin subjects might rely more on the pitch contour when perceiving tones. This view is supported by Chandrasekaran et al. (2007), in which they compared the differences in the acoustic dimensions (pitch height or pitch contour) that native Chinese speakers and native English speakers primarily relied on when perceiving Mandarin tones and found that for pitch contour was much more important for Mandarin-speaking subjects.

5 Limitation

The relatively short nature of the training procedure remains a limitation of the current study, which may capture long-term learning outcomes to an limited extent. To address this problem, we are now conducting a new experiment with an extended training procedure in order to better assess the retention of the training effect.

6 Conclusion

We trained Mandarin speakers to learn Cantonese tones through perceptual learning paradigm with visual feedback provided. Mandarin speakers' ability to identify Cantonese tones improved significantly after training, demonstrating the effectiveness of visual information in auditory tone learning. Mandarin speakers spontaneously gave the most attention to the tone letter – the explicit visual information during the training process. Our results emphasized the importance of explicit visual information in auditory perceptual learning.

Acknowledgments

This research was partly supported by a fellowship award from the Research Grants Council of the Hong Kong SAR, China (Project No. PolyU/RFS2122-5H01) and an internal grant from The Hong Kong Polytechnic University (Project No. P0048115).

References

Baills, F., Suárez-González, N., González-Fuente, S., & Prieto, P. (2019). Observing and producing pitch gestures facilitates the learning of Mandarin Chinese tones and words. *Studies in Second Language Acquisition, 41*(1), 33–58. https://doi.org/10.1017/S0272263118000074

Boersma, Paul & Weenink, David (2024). Praat: doing phonetics by computer [Computer program]. Version 6.4.18, retrieved 21 August 2024 from http://www.praat.org/

Chandrasekaran, B., Gandour, J. T., & Krishnan, A. (2007). Neuroplasticity in the processing of pitch dimensions: A multidimensional scaling analysis of the mismatch negativity. *Restorative Neurology and Neuroscience, 25*, 195–210.

Chandrasekaran, B., Sampath, P. D., & Wong, P. (2010). Individual variability in cue-weighting and lexical tone learning. *The Journal of the Acoustical Society of America*, *128*(1), 456-465. https://doi.org/10.1121/1.3445785

Chang, Y. H. S., Yao, Y., & Huang, B. H. (2017). Effects of linguistic experience on the perception of high-variability non-native tones. *The Journal of the Acoustical Society of America, 141*(2), EL120-EL126. https://doi.org/10.1121/1.4976037

Chao, Y.-R. (1930). *A system of tone letters*. Le Maître Phonétique.

Francis, A. L., Clocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics, 36*(2), 268–294. https://doi.org/10.1016/j.wocn.2007.06.005

Godfroid, A., Lin, C., & Ryu, C. (2017). Hearing and Seeing Tone Through Color: An Efficacy Study of Web‐Based, Multimodal Chinese Tone Perception Training. *Language Learning, 67*(4), 819‐857. https://doi.org/10.1111/lang.12246

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

Jongman, A., Qin, Z., Zhang, J., & Sereno, J. A. (2017). Just noticeable differences for pitch direction, height, and slope for Mandarin and English listeners. *The Journal of the Acoustical Society of America, 142*(2), EL163-EL169. https://doi.org/10.1121/1.4995526

Liu, Y., Wang, M., Perfetti, C. A., Brubaker, B., Wu, S., & MacWhinney, B. (2011). Learning a Tonal Language by Attending to the Tone: An In Vivo Experiment. *Language Learning, 61*(4), 1119–1141. https://doi.org/10.1111/j.1467-9922.2011.00673.x

Mayer, R. E. (Ed.). (2001). *Multimedia learning*. Cambridge, UK: Cambridge University Press.

Morett, L. M., & Chang, L.-Y. (2015). Emphasising sound and meaning: Pitch gestures enhance Mandarin lexical tone acquisition. *Language, Cognition and Neuroscience, 30*(3), 347–353. https://doi.org/10.1080/23273798.2014.923105

Morett, L. M., Feiler, J. B., & Getz, L. M. (2022). Elucidating the influences of embodiment and conceptual metaphor on lexical and non-speech tone learning. *Cognition, 222*, 105014. https://doi.org/10.1016/j.cognition.2022.105014

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford, UK: Oxford University Press.

Peng, G. (2006). Temporal and tonal aspects of Chinese syllables: A corpus-based comparative study of Mandarin and Cantonese. In *Journal of Chinese Linguistics* (Vol. 34, Issue 1, pp. 134–154).

So, C. K., & Best, C. T. (2010). Cross-language Perception of Non-native Tonal Contrasts: Effects of Native Phonological and Phonetic Influences. *Language and Speech, 53*(2), 273–293. https://doi.org/10.1177/0023830909357156

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. https://doi.org/10.1121/1.428217

Wei, Y., Jia, L., Gao, F., & Wang, J. (2022). Visual-Auditory Integration and High-Variability Speech Can Facilitate Mandarin Chinese Tone Identification. *Journal of Speech Language and Hearing Research*, *65*(11), 4096–4111. https://doi.org/10.1044/2022_JSLHR-21-00691

Zhang, K., Li, Y., & Peng, G. (2016). Cognitive representation of phonological categories: The evidence from Mandarin speakers' learning of Cantonese tones. *2016 10th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, 1–5. https://doi.org/10.1109/ISCSLP.2016.7918457

Zhen, A., Van Hedger, S., Heald, S., Goldin-Meadow, S., & Tian, X. (2019). Manual directional gestures facilitate cross-modal perceptual learning. *Cognition, 187*, 178–187. https://doi.org/10.1016/j.cognition.2019.03.004