Speech Prosody 2026
26-29 May 2026, Philadelphia, Pennsylvania, USA

Enhancing Cantonese tone learning in Mandarin speakers: The role of visual feedback and individual differences

Yuqin Shu, Mingxi Lu, Gaode Zhang, Ran Tao, Gang Peng[*]

The Hong Kong Polytechnic University, Hong Kong
yuqin.shu@connect.polyu.hk, mingxiii.lu@connect.polyu.hk, gaode-gordon.zhang@connect.polyu.hk, ran.tao@polyu.edu.hk, gang.peng@polyu.edu.hk

Abstract

While a growing body of research has demonstrated the positive role of visual cues in auditory phonetic learning, most studies provided visual cues directly rather than using feedback-based approaches. This study investigates the role of visual feedback in the perceptual learning of Cantonese tones by Mandarin speakers. Thirty native Mandarin speakers (15 female) participated in a four-day training program within one week, where they received trial-by-trial visual-only feedback after identifying auditory stimuli. Tone identification tasks were administered before, immediately after, and one-week post-training to evaluate perceptual improvement, generalization to unfamiliar talkers, and short-term retention. Participants' cognitive abilities, including working memory, pitch sensitivity, and musicality, were also assessed. Using generalized linear mixed-effects models, the results showed significant post-training improvement in Cantonese tone perception, with the effects generalized to novel talkers and retained after one week. Notably, musicality consistently predicted performance across all post-training stages, while working memory supported initial learning gains but showed reduced effects in generalization and retention. Pitch sensitivity had minimal influence overall. These findings underscore the effectiveness of visual feedback-based training in non-native tone learning, suggesting that visual feedback enhances robust and transferable phonetic category formation in adult learners, as modulated by individual differences in perceptual abilities.

Index Terms: visual feedback, perceptual learning, tone learning, Cantonese tones, Mandarin speaker

1. Introduction

Acquiring non-native tonal contrasts can be challenging, even for speakers of tonal languages. This difficulty is often attributed to the influence of first language (L1) experience and the specific characteristics of both the L1 and the target language tone systems [1-3]. To facilitate tone learning, various training paradigms have been developed. Among these, the use of multisensory input, particularly visual information, has received increasing attention.

Visual cues are believed to enhance auditory processing. Dual-coding theory [4] posits that verbal and non-verbal (e.g., visual) information are processed through separate but interconnected cognitive systems, thereby enhancing learning when information is presented in both modalities, as it fosters multiple cognitive connections. Building on this idea, the cognitive theory of multimedia learning [5] integrates cognitive load theory and emphasizes that the interplay between auditory and visual modalities facilitates deeper cognitive processing by reducing extraneous cognitive load.

Prior research has demonstrated that integrating visual cues can significantly improve learners' speech perception and comprehension (e.g., [6-8]). In the context of tone learning, various types of visual cues have been shown to facilitate the perception of tonal contrasts, such as real pitch contours [9], static or dynamic visualization of pitch changes [10-12], hand gestures [13-15], numbers [9, 11], and even colors [11]. Notably, in most of these studies, visual information was provided directly during training, with less emphasis on the role of feedback in the learning process. However, both classroom-based [16] and labortory experimental studies [17] have shown that explicit corrective feedback can lead to greater learning performance in speech perception, which highlights the importance of feedback in perceptual learning. Building on this insight, one of the aims of the present study is to examine whether visual feedback-based training can effectively enhance the perception of non-native tonal contrasts. Additionally, we seek to determine whether learning effects can generalize to untrained talkers and be retained after a short delay.

In addition to training design, previous research has shown that substantial individual differences exist in tone learning outcomes. A growing body of work suggests that learners' cognitive and perceptual abilities play an important role in shaping training success, with musical ability, working memory, and pitch sensitivity receiving the most attention [18–20]. However, most previous studies have focused primarily on overall immediate learning outcomes, with less attention to whether these cognitive abilities contribute differently to immediate learning, generalization to untrained talkers, and retention over time. It therefore remains less known whether their influence remains stable over time or varies across assessment contexts.

While most prior studies have focused on tone learning in non-tonal language learners (e.g., English speakers learning Mandarin), less is known about how individual cognitive abilities modulate the acquisition of a new tone system by speakers of tonal languages. Cantonese provides an ideal case for such investigation. Compared with Mandarin, which has four tones (T55, T35, T214, T51), Cantonese has six unchecked tones, including three level tones (T55, T33, T22), two rising tones (T25, T23), and one low-falling tone (T21). These tones differ not only in pitch height but also in pitch contour and slope. In tone space (e.g., pitch height × pitch slope), Cantonese tones are more densely distributed and partially overlapping, whereas Mandarin tones are relatively more discrete [21]. These differences present unique perceptual challenges for Mandarin speakers attempting to acquire Cantonese tones [22].

Therefore, the present study aims to investigate the effectiveness of a visual feedback-based training paradigm in the perceptual learning of Cantonese tones by Mandarin speakers. Specifically, we address the following research

 10.21437/SpeechProsody.2026-104

questions: (1) Does training with visual feedback enhance Mandarin speakers' perception of Cantonese tones, and, if so, can these effects generalize to novel talkers and be retained over time? (2) How do cognitive abilities (e.g., pitch sensitivity, working memory, and musicality) contribute to learning outcomes, generalization, and retention?

2. Methods

2.1. Participants

Participants were 30 native Mandarin speakers (15F, 15M; ages 18-29) from the Hong Kong Polytechnic University. All the participants were born and raised in northern area of China and reported no prior knowledge of Cantonese. At the time of participation, most participants (25 out of 30) had lived in Hong Kong for no more than three months, while the remaining five had resided there for six months to one year. All reported normal hearing, normal or corrected-to-normal vision, and no history of speech or language disorders. All participants were right-handed and had no background in linguistics or psychology. The study was approved by the ethics committee of Hong Kong Polytechnic University, and participants provided written informed consent.

2.2. Stimuli

The stimuli consisted of 24 Cantonese monosyllabic words, which were derived from four carrier syllables (/fɐn/, /fu/, /ji/, and /sɛ/) with 6 Cantonese tones. Among the 24 words, the six /sɛ/-based words served as context stimuli, presented at the very beginning of each block to provide participants with an overall sense of the talker's pitch range. The remaining 18 words were used as the target stimuli for training and testing.

Four native Cantonese speakers (2F, 2M) were recruited to produce these words. Two of them (1F, 1M) served as trained talkers, whose productions were used as training stimuli. The remaining two speakers (1F, 1M) served as untrained (novel) talkers. Each talker produced every word 10 times in a sound-proof booth. For the training phase, one sample was selected from the 10 repetitions produced by each of the two trained talkers, resulting in 48 training tokens (1 sample × 2 talkers × 24 words). For the stimuli to test the generalization effect in novel talkers, three samples were selected from the 10 repetitions produced by each of the 2 novel talkers, resulting in 144 tokens (3 samples × 2 speakers × 24 words). Figure 1 illustrates the tone chart for all tokens of the stimuli.

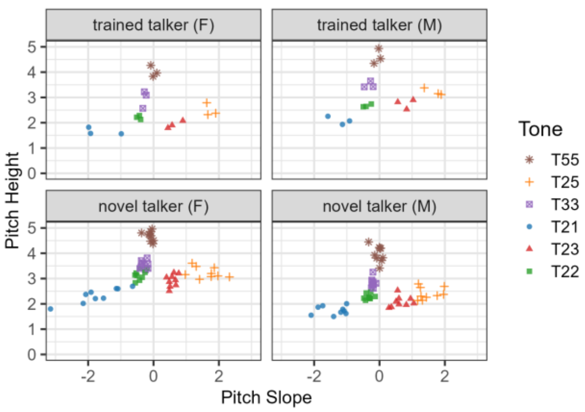

Figure 1: *Tone chart for all stimuli. Pitch slope (x-axis) reflects the slope of the pitch trajectory, and pitch height (y-axis) follows a five-level tone marking system. Each point shows a stimulus token.*

2.3. Procedure

The experiment lasted for two weeks and consisted of three phases: a pre-test, a four-day training phase (completed within one week), and a series of post-training tests. Following training, participants completed post-training tests based on a 2 (Time: immediate vs. delayed) × 2 (Talker: same vs. new) design, yielding four test conditions: post-TT, post-NT, delayed-TT, delayed-NT (referring to trained/novel talkers in immediate post and delayed testing, respectively). All sessions were conducted in a soundproof booth with participants wearing headphones.

In the training sessions, participants completed two blocks (a female talker and a male talker block, counterbalanced across participants) each day. At the beginning of each block, participants listened to a set of context stimuli to let themselves have an impression about the voice and pitch range of each talker, followed by the formal training trials. In each training trial, participants first heard an auditory stimulus, then saw six options displayed on the screen. They were required to select the tone category corresponding to the stimulus by pressing a key. Immediately after their response, visual feedback was provided: a red fixation point indicated an incorrect response, while a blue fixation point indicated a correct one. This was followed by a 3-second display showing the correct visual information associated with the target word. The visual feedback included four types of information: (1) the traditional Chinese character of the word, (2) its English meaning (single word), (3) a number from 1 to 6 indicating the tone category, and (4) Chao tone letters [23], which visually represent pitch changes using a system of vertical and horizontal lines (a right vertical line indicates the overall range of pitch height; a left short bar indicates the starting and ending pitch height and the contour). Prior to the training, participants were informed that Cantonese has six unchecked tones and were explicitly introduced to the meaning of each type of visual cue. The location of each type of visual cues on the screen was randomized across trials. Each day, participants completed 144 training trials (18 words × 1 token per word × 4 repetitions × 2 speaker blocks), which took 20-25 minutes. In total, participants completed 576 training trials across the four-day training phase.

In the testing sessions, tone identification tasks were used to assess participants' perceptual accuracy regarding Cantonese tones. Participants also heard a context at the beginning, followed by formal testing trials. In each trial, participants were presented with an auditory stimulus and six response options on the screen. They were instructed to select the option corresponding to the perceived tone by pressing a key. For test conditions involving the familiar (trained) talkers, each tone identification task included 108 trials (18 words × 3 repetitions × 2 speakers). The same number of trials was used in the unfamilar (novel) talker conditions, comprising 108 trials (18 words × 3 tokens for each word × 1 repetition × 2 speakers). Each task took about 10 minutes.

Participants additionally completed three cognitive tasks during the training sessions. First, a digital span task (both forward and backward) was used to measure their working memory capacity. Second, the Musical Ear Test (MET) [24, 25] was used to assess musicality. In the test, participants heard two musical phrases sequentially and were asked to judge whether two phrases were the same or different. Third, an adaptive pitch test [26, 27] was used to evaluete pitch senstivity. In each trial, participants heard two pure tones and were asked to determine whether the second tone was higher or lower in pitch than the first.

2.4. Data analysis

Tone identification results across all test conditions were submitted to a single generalized mixed-effects model (GLMM) to examine the effects of training and Cantonese tone categories. In addition, separate GLMMs were constructed to investigate the extent to which individual cognitive abilities predicted (1) immediate learning outcomes, (2) generalization to novel talkers, and (3) retention over a one-week delay.

Working memory score was calculated as the average of the z-scored span lengths from both forward and backward digit span tasks. Pitch sensitivity was operationalized as the inverse of the minimal detectable frequency difference (in Hz); the resulting values were then standardized. Musicality scores from the MET were also standardized prior to inclusion in the models.

For each model, a maximal random-effects model was initially built [28] and then simplified to a best-fitting model [29] based on likelihood ration tests, to achieve convergence and improve model stability. All analyses were conducted in R (v4.4.1) [30] using the lme4 package (v1.1-35.5) [31].

3. Results

3.1. Tone identification accuracy across tests

Figure 2(a) depicts the mean accuracy of tone identification tasks in the pre-test, post-test, and delayed test, for both trained talkers and novel talkers. Figure 2(b) illustrates the mean identification accuracy of each tone across test sessions and talkers. A GLMM was constructed with accuracy (0, 1) as the dependent variable. Fixed effects included Test condition (pre, post-TT, post-NT, delayed-TT, delayed-NT), Tone (T55, T25, T33, T21, T23, T22), and their interaction. The random effects included random intercepts for subjects and items, as well as random slopes for Tone by item. Significant effects were found for Test (χ^2 (4) = 53.64, p < .001), indicating that participants' tone identification accuracy significantly improved after training; Tone (χ^2 (5) = 79.04, p < .001), suggesting that some tones were more accurately identified than others across conditions; and the Test × Tone interaction (χ^2 (20) = 265.70, p < .001), indicating that the effect of training varied across different tones.

To further examine tone-specific performance across test conditions, post hoc comparisons were conducted. The results showed that, overall, the accuracy of all tones significantly improved from pretest to all four post-training test conditions (ps < .05) except for T55. For T55, significant improvements were observed only from the pretest to the post-TT and delayed-TT conditions (ps < .05), but not to the post-NT or delayed-NT conditions, likely due to its already high baseline accuracy in the pretest.

In terms of the effects of generalizing to novel talkers and retention for one week, all tones but T21 showed no significant differences across post-TT, post-NT, delayed-TT, and delayed-NT, suggesting successful generalization from familiar to unfamiliar talkers, as well as sustained retention. As for T21, it showed stable retention across time with no significant difference observed from post-TT to delayed-TT or from post-NT to delayed-NT. Moreover, T21's perceptual accuracy was even higher in novel talkers than same talkers, with significant increases observed from post-TT to post-NT (p = .039).

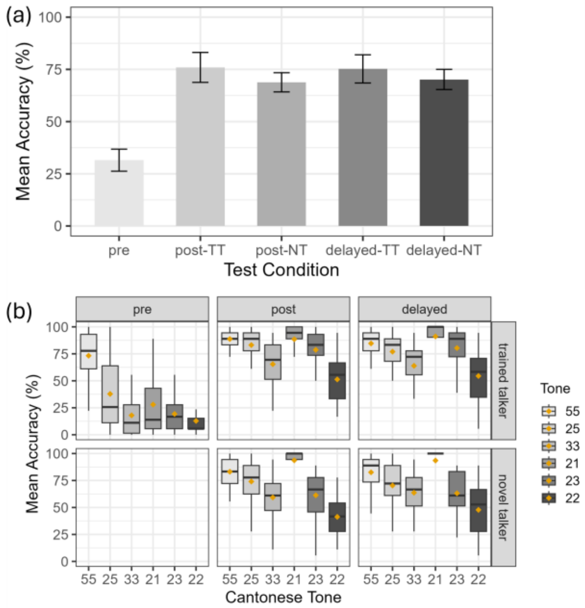

Figure 2: *(a) Overall mean accuracy in the tone identification task across different test sessions. Error bars represent standard deviations. (b) Boxplots showing identification accuracy for the six Cantonese tones across test sessions. Orange diamonds indicate mean accuracy values.*

3.2. Cognitive predictors of Cantonese tone learning through visual feedback

The model evaluating the predictors of immediate learning performance included fixed effects of Time (pre vs. post-TT), the three cognitive scores, and their interactions with Time. The random effects included random intercepts for subjects and items. The results revealed significant interactions between Time and working memory (β = 0.29, SE = 0.09, z = 3.08, p = .002), Time and musicality (β = 0.51, SE = 0.08, z = 6.63, p < .001), indicating that individuals with higher working memory and musical ability showed greater improvement after training. Simple slope analyses further revealed that, musicality significantly predicted post-training performance (β = 0.38, p = .02), but not pretest scores, suggesting a specific contribution to learning outcomes. For working memory, while neither the pretest nor post-test slope was individually significant, the significant interaction with Time suggested that participants with higher working memory benefitted more from training.

The model examining generalization to novel talkers included fixed effects of Talker (trained vs. novel), the cognitive predictors, and their interactions. The random effects included random intercepts for subjects and items. The model revealed a significant effect of musicality (β = 0.34, SE = 0.15, z = 2.27, p = .02), indicating that individuals with higher musical ability performed better overall, regardless of talker familiarity. No significant main effects of working memory or pitch sensitivity were found (ps > .19). The interaction between Talker and working memory was not statistically significant (β

= -0.17, SE = 0.09, z = -1.83, p = .068), nor were the interactions between Talker and musicality or pitch sensitivity (ps > .43).

The model testing retention over time included fixed effects of Time (post-TT vs. delayed-TT), the cognitive factors and their interactions. The random effects inlcuded random intercepts for subjects and items. The retention model revealed a significant main effect of musicality (β = 0.32, SE = 0.16, z = 2.06, p = .039). Simple slope analyses showed that musical aptitude significantly predicted performance at both post-test (β = 0.32, p = .04) and delayed test (β = 0.45, p = .004), suggesting a stronger contribution to long-term retention. In contrast, working memory showed a significant negative interaction with Time (β = -0.20, p = .034), with its predictive effect diminishing over time. Pitch sensitivity was not a significant predictor of retention performance.

4. Discussion

In this study, we examined a visual feedback-based training paradigm for the perceptual learning of Cantonese tones by native Mandarin speakers. We also investigated how individual differences in working memory, musicality, and pitch sensitivity influenced performance across three stages: immediate learning, generalization to novel talkers, and retention. Our results showed that participants significantly improved their identification of the six Cantonese tones following training, although accuracy varied across tones. The training also led to reliable generalization to novel talkers and retention over time. We further found that across stages, musicality emerged as a consistent and robust predictor of performance. Individuals with higher musical aptitude showed greater immediate learning gains, better generalization, and stronger retention, with an even stronger association observed at the delayed test. In contrast, working memory predicted immediate learning outcomes but did not reliably predict generalization or retention. Pitch sensitivity showed limited predictive value across all three stages of testing.

These findings demonstrate that training programs incorporating visual feedback are efficient for tone learning, with benefits that generalize across talkers and retain over time. We think that this may benefit from both the use of corrective feedback and the visual-only information display. First, corrective feedback may enhance the learner’s internal representation of tone categories: correct feedback reinforces accurate tone-label mappings and may elicit positive reinforcement, while incorrect feedback provides opportunities for error correction. Second, during visual cues presentation, participants did not hear the target stimulus again but were instead prompted to mentally recall and re-evaluate the tone based solely on visual input. This process may require additional cognitive effort, but it also encourages learners to focus on abstract tonal features rather than memorizing specific acoustic parameters. Over time, this may help learners develop more flexible and robust perceptual mappings, which could explain the observed generalization and retention effects. The finding that working memory was a significant moderator of immediate learning outcome, but not of generalization or retention, may further support this interpretation. Working memory may contribute primarily during the active learning phase, when learners must integrate auditory input, visual feedback, and internal representations to form new tone categories. Its limited role in retention suggests that such cognitive resources are less critical once tonal representations have been established. The diminished effect of working memory in the generalization task may reflect increased difficulty in mapping familiar tone categories onto the acoustically variable input from a new talker. When the acoustic cues deviate from the trained input, working memory may be less effective in supporting tone identification.

Our findings suggest that musicality is a robust predictor of Cantonese tone learning in Mandarin-speaking learners. It significantly predicted not only immediate learning outcomes, but also showed main effects in both generalization and retention, with its influence marginally increasing at the delayed test. This pattern aligns with previous research showing that musicality facilitates tone learning [20]. In the context of the current training paradigm, musicality may have contributed in two ways: first, by enabling more fine-grained perception of pitch patterns during auditory encoding; and second, by supporting audiovisual integration during training. Some participants with higher musicality in our sample had several years of amateur music training, which may have enhanced their ability to associate explicit visual cues, such as tone letters, with corresponding auditory features.

In contrast, pitch sensitivity did not significantly predict performance at any stage of learning, and only showed a marginal effect on immediate learning gains. This diverges from some prior findings in non-tonal language learners [19], where pitch discrimination ability has been linked to tone learning outcomes. One possible explanation is that our pitch sensitivity measure assessed non-speech processing of pure tones, which may not directly translate to speech-based pitch perception—reflecting a dissociation between speech and non-speech pitch processing. Another possibility is that, as native speakers of a tone language, Mandarin learners already possess well-developed, automatized tone processing abilities, rendering basic pitch sensitivity less relevant in predicting tone learning success.

It is important to note that the observed training effects should be understood as the outcome of an integrated instructional approach. Each component of the present design, including the presence of visual information, corrective feedback, repeated perceptual training, and explicit instruction of Cantonese tones, may have contributed to the overall learning gains. Rather than examining these elements independently, the current study aimed to evaluate the effectiveness of a visual feedback-based training paradigm as a whole. The specific contribution of each component warrants further investigation through controlled comparisons.

To summarize, this study demonstrates that visual feedback-based training effectively enhances Mandarin speakers’ perceptual learning of Cantonese tones, yielding not only immediate improvement but also robust generalization and sustained retention. Among individual learner factors, musical ability consistently predicted success across all stages, while working memory supported initial learning but showed limited long-term effects, and pitch sensitivity played only a marginal role. These findings highlight how individual differences interact with training design to shape cross-language tone learning, and underscore the potential of multimodal feedback in promoting phonetic learning beyond the native language.

5. Acknowledgements

This research is supported by a fellowship award from the Research Grants Council of the Hong Kong SAR, China (Project No. PolyU/RFS2122-5H01).

6. References

[1] C. K. So and C. T. Best, “Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences,” *Language and Speech,* Article vol. 53, pp. 273-293, 2010.

[2] A. L. Francis, V. Clocca, L. Ma, and K. Fenn, “Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers,” *Journal of Phonetics,* Article vol. 36, no. 2, pp. 268-294, 2008.

[3] Y.-C. Hao, “Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers,” *Journal of Phonetics,* Article vol. 40, no. 2, pp. 269-279, 2012.

[4] A. Paivio, *Mental representations: A dual coding approach*. Oxford university press, 1986.

[5] R. E. Mayer, “Multimedia learning,” in *Psychology of Learning and Motivation*, vol. 41: Academic Press, 2002, pp. 85-139.

[6] V. Hazan, A. Sennema, A. Faulkner, M. Ortega-Llebaria, M. Iba, and H. Chung, “The use of visual cues in the perception of non-native consonant contrasts,” *Journal of the Acoustical Society of America,* vol. 119, no. 3, pp. 1740-1751, 2006.

[7] S. Kawase, B. Hannah, and Y. Wang, “The influence of visual speech information on the intelligibility of English consonants produced by non-native speakers,” *Journal of the Acoustical Society of America,* Article vol. 136, no. 3, pp. 1352-1362, 2014.

[8] L. E. Bernstein, E. T. Auer, J. Jiang, and S. P. Eberhardt, “Auditory perceptual learning for speech perception can be enhanced by audiovisual training,” *Frontiers in Neuroscience,* Original Research vol. Volume 7, 2013.

[9] Y. Liu, M. Wang, C. A. Perfetti, B. Brubaker, S. Wu, and B. MacWhinney, “Learning a tonal language by attending to the tone: An in vivo experiment,” *Language Learning,* vol. 61, no. 4, pp. 1119-1141, 2011.

[10] Y. Wei, L. Jia, F. Gao, and J. Wang, “Visual-auditory integration and high-variability speech can facilitate Mandarin Chinese tone identification,” *Journal of Speech, Language and Hearing Research,* Article vol. 65, no. 11, pp. 4096-4111, 2022.

[11] A. Godfroid, C.-H. Lin, and C. Ryu, “Hearing and seeing tone through color: An efficacy study of web-based, multimodal Chinese tone perception training,” *Language Learning,* vol. 67, no. 4, pp. 819-857, 2017.

[12] A. Zhen, S. Van Hedger, S. Heald, S. Goldin-Meadow, and X. Tian, “Manual directional gestures facilitate cross-modal perceptual learning,” *Cognition,* Article vol. 187, pp. 178-187, 2019.

[13] L. M. Morett, J. B. Feiler, and L. M. Getz, “Elucidating the influences of embodiment and conceptual metaphor on lexical and non-speech tone learning,” *Cognition,* vol. 222, p. 105014, 2022.

[14] L. M. Morett and L.-Y. Chang, “Emphasising sound and meaning: pitch gestures enhance Mandarin lexical tone acquisition,” *Language, Cognition and Neuroscience,* Article vol. 30, no. 3, pp. 347-353, 2015.

[15] F. Baills, N. Suarez-Gonzalez, S. Gonzalez-Fuente, and P. Prieto, “Observing and producing pitch gestures facilitates the learning of Mandarin Chinese tones and words,” *Studies in Second Language Acquisition,* Article vol. 41, no. 1, pp. 33-58, 2019.

[16] A. H. Lee and R. Lyster, “The effects of corrective feedback on instructed L2 speech perception,” *Studies in Second Language Acquisition,* vol. 38, no. 1, pp. 35-64, 2016.

[17] M. I. Lehet, K. M. Fenn, and H. C. Nusbaum, “Shaping perceptual learning of synthetic speech through feedback,” *Psychonomic Bulletin & Review,* vol. 27, no. 5, pp. 1043-1051, 2020.

[18] T. K. Perrachione, J. Lee, L. Y. Y. Ha, and P. C. M. Wong, “Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design,” *Journal of the Acoustical Society of America,* Article vol. 130, no. 1, pp. 461-472, 2011.

[19] A. R. Bowles, C. B. Chang, and V. P. Karuzis, “Pitch ability as an aptitude for tone learning,” *Language Learning,* Article vol. 66, no. 4, pp. 774-808, 2016.

[20] P. C. M. Wong and T. K. Perrachione, “Learning pitch patterns in lexical identification by native English-speaking adults,” *Applied Psycholinguistics,* Article vol. 28, no. 4, pp. 565-585, 2007.

[21] G. Peng, “Temporal and tonal aspects of Chinese syllables: A corpus-based comparative study of Mandarin and Cantonese,” *Journal of Chinese Linguistics,* Article vol. 34, no. 1, pp. 134-154, 2006.

[22] K. Zhang, G. Peng, Y. Li, J. W. Minett, and W. S. Y. Wang, “The effect of speech variability on tonal language speakers' second language lexical tone learning,” *Frontiers in Psychology,* Article vol. 9, 2018.

[23] Y.-R. Chao, “A system of tone letters,” *Le maître phonétique,* 1930.

[24] M. Wallentin, A. H. Nielsen, M. Friis-Olivarius, C. Vuust, and P. Vuust, “The Musical Ear Test, a new reliable test for measuring musical competence,” *Learning and Individual Differences,* vol. 20, no. 3, pp. 188-196, 2010.

[25] A. I. Correia, M. Vincenzi, P. Vanzella, A. P. Pinheiro, C. F. Lima, and E. G. Schellenberg, “Can musical ability be tested online?,” *Behavior Research Methods,* Article vol. 54, no. 2, pp. 955-969, 2022.

[26] S. Wiener and J. Liu, “Effects of perceptual abilities and lexical knowledge on the phonetic categorization of second language speech,” *JASA Express Letters,* vol. 1, no. 4, p. 045202, 2021.

[27] J. Mandell. “Adaptive pitch test.” http://tonometric.com (accessed November 5, 2025).

[28] D. J. Barr, R. Levy, C. Scheepers, and H. J. Tily, “Random effects structure for confirmatory hypothesis testing: Keep it maximal,” *Journal of Memory and Language,* vol. 68, no. 3, pp. 255-278, 2013.

[29] D. Bates, R. Kliegl, S. Vasishth, and H. Baayen, “Parsimonious mixed models,” *arXiv preprint arXiv:1506.04967,* 2015.

[30] R Developmental Core Team, “*R: A language and environment for statistical computing.*” Vienna, Autria, 2025.

[31] D. Bates, M. Mächler, B. Bolker, and S. Walker, “Fitting linear mixed-effects models using lme4,” *Journal of Statistical Software,* vol. 67, no. 1, pp. 1 - 48, 2015.