Interspeech 2018
2-6 September 2018, Hyderabad

Experience-dependent influence of music and language on lexical pitch learning is not additive

Akshay Raj Maggu[1], *Patrick C. M. Wong*[1], *Hanjun Liu*[2], *Francis C. K. Wong*[3]

[1]The Chinese University of Hong Kong, Hong Kong SAR
[2]Sun Yat-sen University, China
[3]Nanyang Technological University, Singapore

akshay@cuhk.edu.hk, p.wong@cuhk.edu.hk, lhanjun@mail.sysu.edu.cn, franciswong@ntu.edu.sg

Abstract

Research studies provide evidence for the facilitative effects of musical and linguistic experience on lexical pitch learning. However, the effect of interaction of linguistic and musical pitch experience on lexical pitch processing is a matter of ongoing research. In the current study, we sought to examine the effect of combination of musical and linguistic pitch experience on learning of novel lexical pitch. Using a 10-session pseudoword-picture association training paradigm, we compared the learning performance of musicians and non-musicians who either spoke a non-tone language, spoke one tone language, or spoke two tone languages. Among the non-tone language speakers, we found that musicians showed enhanced learning of novel lexical pitch as compared to non-musicians. In comparison, among the tone-language speakers, we found no significant difference in the learning performance of musicians and non-musicians no matter they spoke one or more tone languages. We conclude that though musical experience facilitates linguistic pitch learning, the effects of combination of musical and linguistic pitch experience are not additive i.e. possessing both types of pitch experience is no better than possessing either one of them and knowing two tone languages does not facilitate the learning of a new tone language beyond the knowledge of one.

Index Terms: language-music interaction, speech perception, training, lexical tone

1. Introduction

Previous studies [1]–[6] revealed a facilitatory effect of linguistic and musical pitch experience towards linguistic perception and learning. However, how the interaction of effects of linguistic and musical pitch experience affect learning of novel lexical tone system is an intriguing question. Recently, it has been reported that the facilitation effect of musical experience on linguistic pitch learning is not a straightforward but a complex phenomenon [7]. In order to further understand the complexity of interaction of linguistic and musical experience, in the current study, we compared the learning performance of the individuals with varied musical and/or linguistic experience towards novel lexical pitch learning.

In the past, studies have been conducted using speakers of non-tone and tone languages to understand the facilitative effect of musical experience on linguistic perception. It has been reported that English musicians are better than English non-musicians on lexical tone perception and learning [5], [6], [8]. However, in tone language speakers with musical experience, researchers [9]–[11] have reported a lack of correlation between language and music perception. Further, Cooper and Wang [7] compared the performance of Thai and English musicians and non-musicians for learning novel linguistic pitch contours. Consistent with the previous studies [6], they found that English musicians learned the novel lexical pitch more accurately than English non-musicians. However, they found no significant difference between Thai musicians and English musicians on learning performance. Further, Thai musicians showed slightly poorer performance than Thai non-musicians.

In order to further understand the complexity of effects of combination of linguistic and musical experience, in the current study, musicians and non-musicians who either spoke a non-tone language (English), spoke one tone language (Mandarin), and those who spoke two tone languages (Cantonese and Mandarin) were compared on their learning performance of a novel lexical tone system using a 10-session pseudoword-picture association task. In this task, all the subjects learned to associate pseudowords with novel lexical tones with pictures of objects presented on a computer screen [6]. To successfully learn the pseudoword-picture pairings, participants had to be sensitive not just to the segmental features, but also to the novel suprasegmental features, namely, changes in pitch pattern within syllables that were not used in their native languages. If the facilitative effect on lexical pitch processing by the two types of pitch experience is additive, it is predicted that Cantonese-Mandarin musicians would perform the best followed by Cantonese-Mandarin non-musicians and Mandarin musicians followed by Mandarin non-musicians and English musicians, and finally English non-musicians should perform the worst. However, if the effects are not additive but complex, we should be able to see a different hierarchy of effects that could range from no facilitative effects to adverse effects of combination of multiple experiences.

2. Method

A 2-by-3 between-subjects design was employed with musical experience (musician vs. non-musician) and language background (English vs. Cantonese-Mandarin vs. Mandarin) as the between-subjects factors.

2.1. Participants

A total of 70 subjects participated in the study. Among the monolingual English speakers, 13 of them were musicians and 13 of them were non-musicians. The Cantonese-Mandarin group (7 musicians and 12 non-musicians) spoke Cantonese as

 10.21437/Interspeech.2018-2104

their first language, and Mandarin as their second language. As for the Mandarin subjects, 10 of them were musicians and 15 of them were non-musicians. Neither the Cantonese-Mandarin nor the Mandarin subjects spoke another Chinese dialect. Musicians were defined as participants with 6 or more years of formal musical training on any musical instrument while non-musicians were defined as those with less than 3 years of formal musical training. All native English-speaking participants were recruited from Northwestern University, Illinois, USA. Chinese-speaking non-musicians were recruited from Sun Yat-sen University, Guangdong, China. Chinese-speaking musicians were recruited from Xinghai Conservatory of Music, Guangzhou, China, and from the Department of Music at South China Agricultural University, Guangzhou, China. All the participants had provided their written consent before participation and the procedures were approved by the Institutional Review Board at the respective sites, Northwestern University, Sun Yat-sen University and South China Agricultural University.

2.2. Stimuli

The stimuli consisted of Thai lexical tones. Thai language has five linguistic pitch contours: a falling tone, a high tone, a low tone, a middle tone, and a rising tone (Figure 1). The five pitch contours of the Thai language were each superimposed onto three English pseudo-syllables ([pæʃ], [nik], and [fæs]), using the pitch synchronous overlap-add method in Praat. This created 15 pseudowords in total. Stimulus output was set at a comfortable listening level.

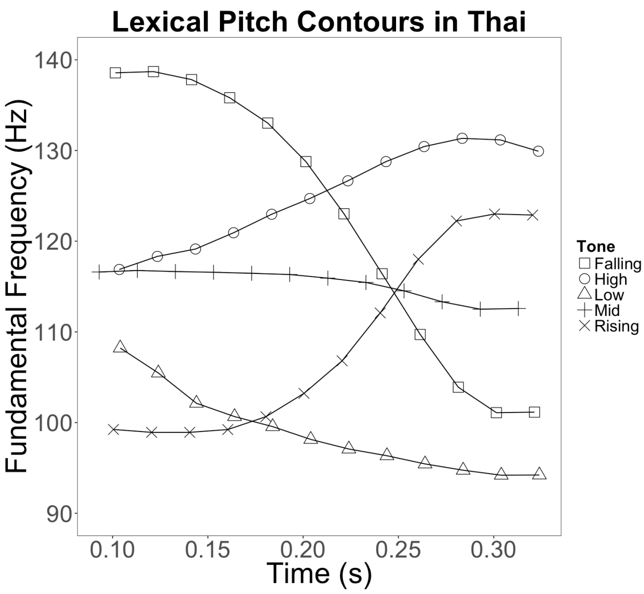

Figure 1: *F0 contours of the five Thai lexical tones (F0 ranges: Falling: 102-138 Hz, High: 118-132 Hz, Mid: 115-118 Hz, Low: 92-108 Hz, Rising: 100-124 Hz)*

2.3. Procedure

The sound-to-word learning program used in the current study has been described in detail by Chandrasekaran et al. [12] and was similar to the previous studies [6], [13], [14]. Each of the 15 spoken pseudowords was paired with an object, visually represented by a unique cartoon picture. Participants took part in 30-45 minute training sessions for 10 days with no more than two days gap between each session. Participants were trained to associate each pseudoword with its corresponding object. During each training session, pseudowords were presented in three blocks via supra-aural headphones, with the corresponding object presented visually on a computer screen. Each block consisted of a single pseudo-syllable, differing only in lexical pitch. Each block consisted of six presentations of each of the five pseudowords corresponding to the pseudo-syllable (30 presentations in total) in pseudo-random order, separated by 3 sec of silence. After each training block, participants were tested on the five pseudowords they had just heard (see Figure 2 for an example). Each pseudoword was presented in a pseudo-randomized sequence, and participants were instructed to select the corresponding object from the five options displayed visually via the computer monitor. If participants selected the incorrect object, the correct object was highlighted prior to the next pseudoword being presented.

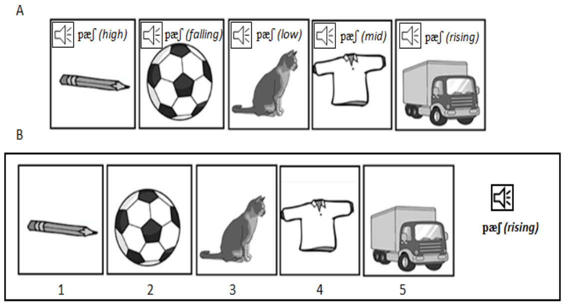

Figure 2: *An example of the training and testing paradigm: (A) In a training session, subjects learned to associate pseudowords with pictures; (B) In the testing phase, they were asked to identify the picture in response to the pseudoword that they had learned during the training sessions.*

After the three training blocks, participants took part in the test phase of the session. Each of the 15 pseudowords was again presented in a pseudo-randomized sequence, and participants were instructed to select the corresponding object from the 15 options displayed visually via the computer monitor. No feedback was provided during the test phase, and as we were solely interested in the accuracy scores and not reaction times, no response time limit was imposed.

3. Results

Figure 3 reveals the average learning curves of the subject groups across the ten sessions of training. Overall, we found that among the musicians, non-tone language speakers, Mandarin and Cantonese-Mandarin speakers had comparable learning curves (Figure 3(A)) while among the non-musicians, non-tone language speakers showed a shallower learning curve compared to Mandarin and Cantonese-Mandarin speakers (Figure 3(B)).

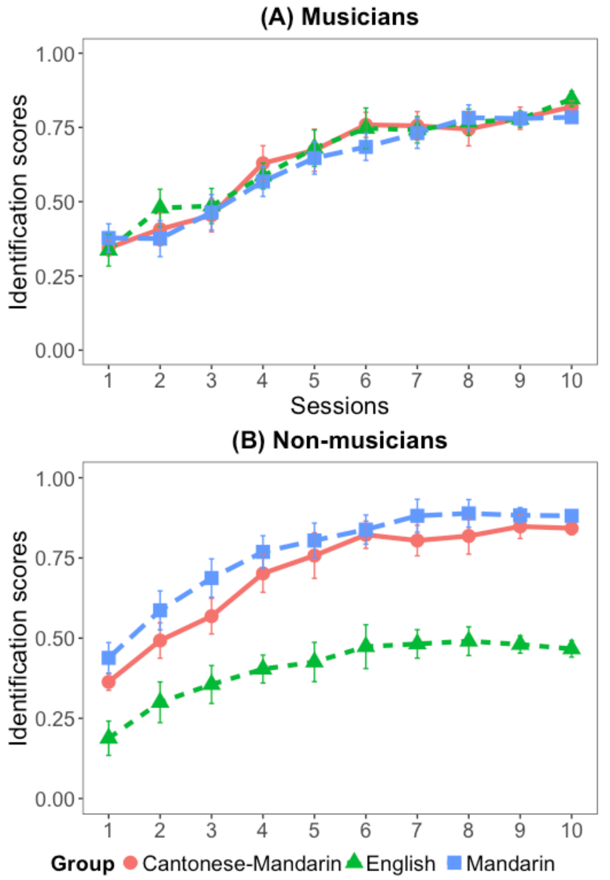

Figure 3: *Comparison of non-tone monolinguals (English), Mandarin, and Cantonese-Mandarin speakers on the learning curves of tone-word identification for the (A) Musician; and (B) Non-musician groups. Error bars = ± SEM*

The word identification (word ID) scores, in terms of percentage correct, in the final session (session 10) are presented in Figure 4. Overall, the mean word ID score across all participants was 77.11% (SD = 20.25%). A 2 (Musical Experience) × 3 (Language Background) ANOVA revealed main effects of both Musical Experience ($F(1, 64) = 6.01$, $p < .05$, $\eta_p^2 = .087$) and Language Background ($F(2, 64) = 12.27$, $p < .001$, $\eta_p^2 = .277$), and a significant interaction effect of the two factors ($F(2, 64) = 20.05$, $p < .001$, $\eta_p^2 = .385$). A Simple-Effect Analysis was conducted to investigate the nature of the interaction, specifically to look at the effect of Language Background on musicians and non-musicians. For musicians, the effect of Language Background was not significant ($F(2, 64) = 0.53$, $p = .589$, $\eta_p^2 =.016$). In the other words, English musicians, Cantonese-Mandarin musicians and Mandarin musicians did not perform differently in the word identification task. As for the non-musicians, the effect of Language Background was significant ($F(2, 64) = 34.68$, $p < .001$, $\eta_p^2 = .520$). Post-hoc pairwise comparisons revealed that both the Cantonese-Mandarin non-musicians and the Mandarin non-musicians outperformed the English non-musicians (both $p < .001$, Bonferroni corrected).

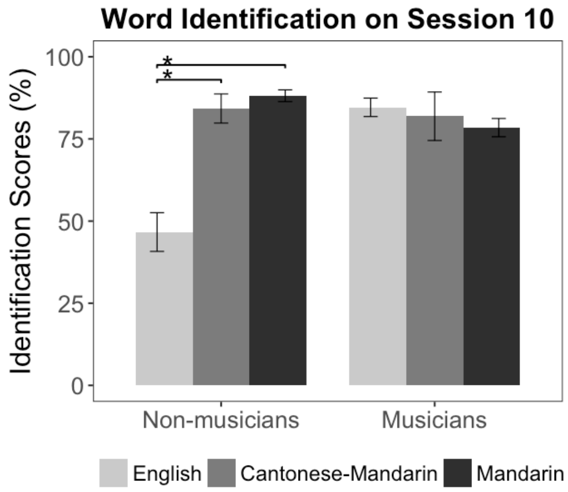

Figure 4: *Word identification scores (percentage correct) on the final training session for the musicians and non-musicians across the three language backgrounds: non-tone-language (English speakers), Cantonese-Mandarin speakers and Mandarin speakers. The nature of the significant interaction effect (Musical Experience by Language Background) was driven by the significant differences between 1) Cantonese-Mandarin non-musicians and English non-musicians; and 2) Mandarin non-musicians and English non-musicians (*p < .001; Error bars = ± SEM).*

Results from the current study reveal how musical pitch experience and linguistic pitch experience may interact and influence the learning of a new tonal system. More specifically, we tested the hypothesis whether multiple pitch experiences are additive. The evidence of this comes from two sources. Firstly, regardless of language background, all musicians performed equally in the word identification task. This suggests that experience with a tone language does not provide additional advantage on top of musical pitch experience in learning a new tone system. Secondly, although non-musicians who speak a tone language (the Mandarin group) outperform English-speaking non-musicians while speaking an additional tone language (the Cantonese-Mandarin group) does not provide additional advantage (Figure 4).

4. Discussion

In the current study, we investigated the interactive effects of varied degrees of musical and/or tone language experience on lexical and musical pitch perception at the behavioral level. The main finding of our study is that though musical experience facilitates perception of lexical tones, the interactive effects of language and music are not additive, but complex. We found that Cantonese-Mandarin musicians (experience with two tone languages and music) performed at the same level of identification accuracy as Cantonese-Mandarin non-musicians (with experience in two tone languages but not music), tone (Mandarin) non-musicians and non-tone (English) musicians, who performed better than non-tone (English) non-musicians. Furthermore, we found that Mandarin musicians with combined language and musical experience trended towards performing slightly less accurately than Mandarin non-musicians.

For the non-tone language speakers, the current findings revealed a facilitation effect of music on lexical tone perception, consistent with the previous findings [1], [2], [5]–[8]. English musicians clearly outperformed the English non-musicians on the tone-word identification training. These findings also stand consistent with Wong and Perrachione [6] and Cooper and Wang [7] who illustrated the enhancing effect of musical experience for learning of novel linguistic pitch. Unlike previous studies [1], [2], the listeners were not asked to simply identify different phonemes, but they were also asked to situate the phonemic contrasts at the word level to differentiate meaning. Surprisingly, Mandarin musicians performed slightly worse than the Mandarin non-musicians on the tone-word identification measures. These findings are consistent with Cooper and Wang [7] who also found that the Thai musicians performed poorly than Thai non-musicians on a tone-word identification-training paradigm. One might speculate that these findings stem from different mechanisms involved in learning a tone language versus learning music. Additionally, our findings revealed that listeners with multiple tone language and musical experience performed similar to listeners with one tone language experience (Mandarin), and non-tone language (English) musicians, on tone-word identification task, suggestive of lack of additivity of experience-dependent effects.

5. Conclusion and future directions

From the current findings, we can conclude that the effect of combination of different types of experiences (language and/or music) are not simply additive towards novel lexical pitch learning. Future studies could consider investigating this research question from a neurophysiological standpoint by using auditory event-related potentials such as frequency following response [15]–[17] and/or cortical pitch response [18].

6. Acknowledgements

This research was supported by the Dr Stanley Ho Medical Development Foundation and National Institutes of Health (USA) Grant No. R01DC008333 to P.C.M.W. and Singapore Ministry of Education Academic Research Fund Tier 1 Grant No. RG72/17 to F.C.K.W.

7. References

[1] J. A. Alexander, P. C. M. Wong, and A. R. Bradlow, “Lexical tone perception in musicians and non-musicians.,” in *Interspeech*, 2005, pp. 397–400.

[2] T. L. Gottfried, “Music and language learning,” *Lang. Exp. Second Lang. Speech Learn.*, pp. 221–237, 2007.

[3] T. L. Gottfried, A. M. Staby, and C. J. Ziemer, “Musical experience and Mandarin tone discrimination and imitation,” *J. Acoust. Soc. Am.*, vol. 115, no. 5, pp. 2545–2545, 2004.

[4] T. L. Gottfried and D. Riester, “Relation of pitch glide perception and Mandarin tone identification,” *J. Acoust. Soc. Am.*, vol. 108, no. 5, p. 2604, 2000.

[5] C.-Y. Lee and T.-H. Hung, “Identification of Mandarin tones by English-speaking musicians and nonmusicians,” *J. Acoust. Soc. Am.*, vol. 124, no. 5, pp. 3235–3248, Nov. 2008.

[6] P. C. M. Wong and T. K. Perrachione, “Learning pitch patterns in lexical identification by native English-speaking adults,” *Appl. Psycholinguist.*, vol. 28, no. 04, pp. 565–585, 2007.

[7] A. Cooper and Y. Wang, “The influence of linguistic and musical experience on Cantonese word learning,” *J. Acoust. Soc. Am.*, vol. 131, no. 6, pp. 4756–4769, 2012.

[8] C.-Y. Lee, A. Lekich, and Y. Zhang, “Perception of pitch height in lexical and musical tones by English-speaking musicians and nonmusiciansa),” *J. Acoust. Soc. Am.*, vol. 135, no. 3, pp. 1607–1615, 2014.

[9] C.-Y. Lee, Y.-F. Lee, and C.-L. Shr, “Perception of musical and lexical tones by Taiwanese-speaking musicians,” *J. Acoust. Soc. Am.*, vol. 130, no. 1, pp. 526–535, 2011.

[10] C.-Y. Lee and Y.-F. Lee, “Perception of musical pitch and lexical tones by Mandarin-speaking musicians,” *J. Acoust. Soc. Am.*, vol. 127, no. 1, pp. 481–490, 2010.

[11] A. Chen, L. Liu, and R. Kager, “Cross-domain correlation in pitch perception, the influence of native language,” *Lang. Cogn. Neurosci.*, vol. 31, no. 6, pp. 751–760, 2016.

[12] B. Chandrasekaran, P. D. Sampath, and P. C. Wong, “Individual variability in cue-weighting and lexical tone learning,” *J. Acoust. Soc. Am.*, vol. 128, no. 1, pp. 456–465, 2010.

[13] F. C. Wong, B. Chandrasekaran, K. Garibaldi, and P. C. Wong, “White matter anisotropy in the ventral language pathway predicts sound-to-word learning success,” *J. Neurosci.*, vol. 31, no. 24, pp. 8780–8785, 2011.

[14] Z. Chen *et al.*, “Transfer effect of speech-sound learning on auditory-motor processing of perceived vocal pitch errors,” *Sci. Rep.*, vol. 5, p. 13134, 2015.

[15] P. C. M. Wong, E. Skoe, N. M. Russo, T. Dees, and N. Kraus, “Musical experience shapes human brainstem encoding of linguistic pitch patterns,” *Nat. Neurosci.*, vol. 10, no. 4, pp. 420–422, Apr. 2007.

[16] A. R. Maggu, F. Liu, M. Antoniou, and P. C. M. Wong, “Neural Correlates of Indicators of Sound Change in Cantonese: Evidence from Cortical and Subcortical Processes,” *Front. Hum. Neurosci.*, vol. 10, 2016.

[17] A. R. Maggu, P. C. Wong, M. Antoniou, O. Bones, H. Liu, and F. C. Wong, “Effects of combination of linguistic and musical pitch experience on subcortical pitch encoding,” *J. Neurolinguistics*, vol. 47, pp. 145–155, 2018.

[18] A. Krishnan, J. T. Gandour, and C. H. Suresh, “Cortical pitch response components show differential sensitivity to native and nonnative pitch contours,” *Brain Lang.*, vol. 138, pp. 51–60, 2014.