



Modulation of Musical Experience and Prosodic Complexity on Lexical Pitch Learning

Xiuli Tong¹, Yee Ching Tang¹

¹Division of Speech and Hearing Sciences, The University of Hong Kong, Hong Kong

xltong@hku.hk, registangyeeching@gmail.com

Abstract

With a non-native (i.e., Thai) pitch-word learning task, the current study examined the impacts of prosodic complexity and musical experience on non-native tone identification and tone word learning by comparing musicians and non-musicians whose native languages exhibit different prosodic complexity, such as Cantonese, Mandarin, and English. We found that for the pre-training tone identification task, musicians outperformed non-musicians, regardless of their native language background. For the tone word learning task, Cantonese musicians outperformed English musicians at the beginning stage of tone word learning. No significant differences were found among non-musicians in the three languages. However, both Cantonese and Mandarin non-musicians outperformed English non-musicians in the final stage of learning, yet there was no difference between musicians. These findings underscore that prosodic complexity and musical experience have dynamic roles in influencing tone identification and tone word learning at different stages.

Index Terms: musical experience, prosodic complexity, tone identification, lexical pitch learning

1. Introduction

Linguistic and extra-linguistic factors have been found to influence second language (L2) learning [1]. One of the linguistic factors, the proximity between native language (L1) and L2 phonetic inventory, plays a key role in determining L2 word learning performance [2]. The extra-linguistic factors, particularly musical experience, have been found to facilitate the perception and production of L2 sound structure [3]. However, none of these previous studies have investigated the impact of prosodic complexity, referring to the density of lexical tone space, and musical experience on lexical pitch learning, which is learning the meaning of lexical words distinguished by phonemic tones. Thus, the goal of the present study is to investigate how prosodic complexity and musical experience modulate lexical pitch learning.

1.1. Effect of prosodic complexity

Studies have demonstrated that tonal language speakers had better imitative pitch production than English speakers [4]. Advantage has also been found in non-native tone identification for subjects with tonal language background over non-tonal language background [5], [6].

Prosodic complexity is related to the density of lexical tone space in a language. According to Barry and Blamey [7], tone space is illustrated as plots of F0 offglide \times F0 onset. The tone spaces across Mandarin, Thai, and Cantonese are shown to be

similar, but they contain four, five, and six contrastive lexical tones respectively. Mandarin has four lexical tones, i.e., high level, high-rising, low-dipping, high-falling tones [8]. Cantonese has six distinctive lexical tones, including three level (high, mid, low) and three contour (high-rising, low-falling, low-rising) tones [9]. Thai has five lexical tones, including three level (high, mid, low) and two contour (falling, rising) tones [10]. Since Cantonese has the most lexical tones, the degree of tonal crowding in its tone space (prosodic complexity) is the highest, while Mandarin has the least dense tone space and prosodic complexity (see Figure 1a, 1b, and 1c). Given the difference of tone space, we hypothesized that speakers with L1 of higher prosodic complexity (Cantonese), perform better than ones with L1 of lower prosodic complexity (Mandarin) in learning non-native tonal languages (Thai).

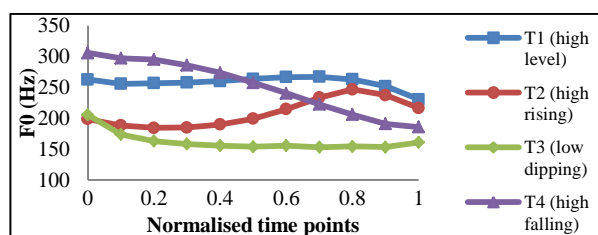


Figure 1a: Tone diagram showing the syllable /fu/ in four Mandarin lexical tones.

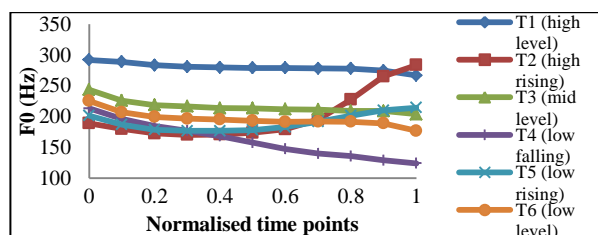


Figure 1b: Tone diagram showing the syllable /fu/ in six Cantonese lexical tones.

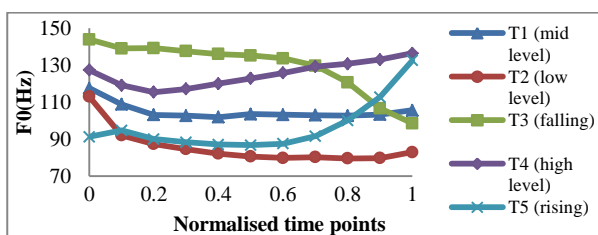


Figure 1c: Tone diagram showing the syllable /te^h ɔ:/ in five Thai lexical tones.

1.2. Effect of musical experience

Musical experience has been shown to have significant influence on lexical tone and speech perception due to the similar acoustic parameters, especially frequency and duration, shared between music and speech [1]. A significant number of studies have revealed that musical experience benefits the perception of non-native lexical tones for non-tonal language participants [11], [12]. Musical experience has also been found to facilitate L2 sound perception and production [3], and tone word learning [13], [14]. In other words, there is substantial evidence revealing that musical experience enhances auditory acuity and is important for tone word learning.

1.3. The present study

This study investigates the relative contributions of prosodic complexity and musical background to non-native tone perception and tone word learning by testing native tonal (Cantonese, Mandarin) and native non-tonal (English) language participants with and without musical background. The Thai tone system was chosen to be the target pitch-word learning because Thai tone system has five tones, which is between Mandarin tone system (four tones) and Cantonese tone system (six tones).

It was expected that the Cantonese group would perform the best while the English group would perform the worst due to the effect of prosodic complexity; whereas musicians would outperform non-musicians due to musical experience. Together with the additive effect of prosodic complexity and musical experience, the performance of non-native tone word learning was expected to result in a hierarchy: Cantonese musicians > Cantonese non-musicians > Mandarin musicians > Mandarin non-musicians > English musicians > English non-musicians (where “>” denotes “outperforms”). Since tonal awareness was shown to contribute to tone word learning [14], it was thus hypothesized that pre-training tone identification scores would also follow the above hierarchy.

2. Method

2.1. Participants

A total of 82 adults were recruited. There were 28 native Cantonese speakers, of which 14 were non-musicians (mean age = 21 years, SD = 1.44 years) without any musical training, and 14 were musicians (mean age = 21 years, SD = 1.91 years), with an average of 11.07 years of musical training; 28 were native Mandarin speakers, of which 14 were non-musicians (mean age = 22 years, SD = 4.22 years), with an average of 0.32 years of musical training (SD = 0.61), and 14 were musicians (mean age = 21 years, SD = 3.09 years), with an average of 9.29 years of musical training (SD = 3.89); 26 were native English speakers, of which 12 were non-musicians (mean age = 26 years, SD = 6.96 years), with an average of 0.64 years of musical training (SD = 0.84), and 14 were musicians (mean age = 23 years, SD = 3.64 years), with an average of 8.17 years of musical training (SD = 2.17).

Musicians are defined as having at least seven years of continuous musical training on Western instruments and/or have obtained official recognition (e.g. ABRSM) of level 8 or above. Moreover, they continue to play an instrument on a regular basis. In contrast, non-musicians are defined as having no musical training within the previous five years and no more than two years of musical experience [15]. All participants

have reported no previous knowledge of Thai or any other tonal languages aside from their native languages.

2.2. Stimuli

2.2.1. Pre-training tone identification

Four Thai CV monosyllables (i.e., /tɛ^hɔː/, /k^huː/, /t^hiːa/, and /wuːa/) were used to carry five Thai lexical tones (i.e., high, mid, low, falling, and rising), resulting in a total of 20 stimuli. These tonal syllables are non-real words in Cantonese, Mandarin, and English, and they were selected with phonotactical probability being considered. This ensured all tonal syllables contained no semantic meaning in the three languages, thus avoided the confounding effect of lexical competition. The stimuli were produced by two native Thai speakers [one male (i.e. M1) and one female (i.e. F1)].

2.2.2. Tone word learning

The stimuli were produced by four other native Thai speakers [two male (i.e. M2, M3) and two female (i.e. F2, F3)]. Three Thai CV monosyllables containing an initial consonant common across Cantonese, English, Thai, and Mandarin with a Thai vowel (i.e. /fɛːw/, /suːa/ and /krːj/) were chosen and produced in five Thai tones. Common consonants were chosen because unfamiliar phonotactics have been shown to hinder word learning [16]. Each word was given a distinct meaning represented by a picture. Pictures of 10 non-real objects were used to avoid memory loading due to previous knowledge of real objects. The pictures were selected from the study conducted by Verma and Brysbaert [17] with visual complexity and objectness being systematically controlled.

2.3. Procedures

Participants completed all the tasks on computers in a sound-proof booth at the University of Hong Kong over two consecutive days (three hours in total).

2.3.1. Pre-training tone identification task

This task started with a familiarization phase so that participants could familiarize themselves with the five Thai tones. The syllable /tɛ^hɔː/ with the five Thai tones were produced by speaker M1, and participants had to identify the tone of each stimulus. Feedback was given and each stimulus was replayed once. The main tone identification task used the same procedure as the familiarization phase, but with no feedback given. Participants were asked to choose the correct tone diagram (see Figure 1c) for 120 stimuli [2 speakers × 4 syllables (/tɛ^hɔː/, /k^huː/, /t^hiːa/, and /wuːa/) × 5 tones × 3 repetitions].

2.3.2. Tone word learning task

All participants completed four 30-min training sessions over two consecutive days, with two sessions each day. Consistent with Cooper and Wang [13], each session consisted of 5 training blocks, a summary block, a review block, and a session test. During each training block, participants were asked to listen to four randomized repetitions of two words, with a picture representing the meaning of the word being shown simultaneously. The two words in each block had no minimal pairs and contained different syllables with different tones (e.g. /tɛ^hɔː1/ and /k^huː3/). Each training block was concluded with a quiz on the two newly learned words. A

stimulus was played and the participants had to choose the corresponding picture. Feedback was provided and the stimulus was replayed once.

A summary block was administered after the five training blocks. Participants had to listen to all 10 words with the corresponding picture being shown twice. It was then followed by a review block, where two randomized repetitions of all 10 words produced by speakers M1 and F1 were played, and participants were asked to select the corresponding picture from the 10 pictures. Feedback was given and the auditory stimulus was replayed once. A session test was conducted to end each training session. After listening to four randomized repetitions from four speakers of all 10 words, participants were asked to select the corresponding picture from the 10 pictures. There were 40 trials with no feedback provided.

3. Results

3.1. Pre-training tone identification

The means of correct identification scores were submitted to a three-way mixed ANOVA, with tone as within-subject factor, and language and musicality as between-subjects factors. A significant main effect of language [$F(2, 76) = 9.36, p < .001$] was found. *Post hoc* (Bonferroni) analysis revealed that both Cantonese ($p < .001$) and Mandarin ($p < .01$) outperformed English significantly, while there was no significant difference between Cantonese and Mandarin speakers ($p = .075$). There was also a significant main effect of tone [$F(4, 304) = 10.41, p < .001$] and musicality [$F(1, 76) = 26.05, p < .001$]. The two-way interaction tone \times language was significant [$F(8, 304) = 3.83, p < .001$]. The interaction effect between language and musicality was marginally significant [$F(2, 76) = 2.72, p = .072$]. However, there was no significant tone \times musicality [$F(4, 304) = 0.45, p = .772$] nor tone \times musicality \times language interactions [$F(8, 304) = 0.42, p = .908$].

To investigate the effects across all groups, a one-way ANOVA with group as independent variable was conducted. As shown in Figure 2, there was a significant main effect of group [$F(5, 76) = 10.33, p < .001$]. *Post hoc* (Bonferroni) analysis showed that Cantonese musicians had significantly higher scores (80.30%) than all other groups ($ps < .001$), except Mandarin musicians (70.60%). Mandarin musicians also outperformed Cantonese non-musicians (52.44%, $p = .049$) and English non-musicians (44.46%, $p = .001$). No significant differences were found in other group comparisons.

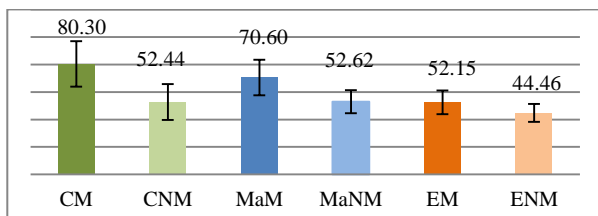


Figure 2: Mean percent correct scores by group in tone identification task. Chance = 20%.

3.2. Tone word learning

Tone word learning performance was determined by the mean percent accuracy scores from the four session tests. To evaluate the overall performance, session test 1 and 4 scores

were submitted to a three-way mixed ANOVA with musicality and language as between-subject factors, and session (session tests 1 and 4) as within-subject factor. There was a significant main effect of session [$F(1, 76) = 198.02, p < .001$], indicating learning process as performance significantly improved after training (from 43.87% to 69.63%). The main effects of language [$F(2, 76) = 22.93, p < .001$] and musicality [$F(1, 76) = 11.03, p = .001$] were also significant. The two-way interaction of session \times language was significant [$F(2, 76) = 4.37, p < .05$], but the two-way interaction of session \times musicality [$F(1, 76) = 0.06, p = .812$] and the three way interaction session \times musicality \times language [$F(2, 76) = 1.63, p = .203$] were not significant.

In order to further examine the interaction between session and language, one-way ANOVAs with language as independent variable were conducted for sessions 1 and 4 respectively. There were significant main effects of languages for sessions 1 [$F(2, 79) = 12.01, p < .001$] and 4 [$F(2, 79) = 25.48, p < .001$]. The results are illustrated in Figure 3. *Post hoc* (Bonferroni) analysis showed that only Cantonese outperformed Mandarin ($p < .05$) and English ($p < .001$) in session 1, whereas both Cantonese and Mandarin were able to outperform English in session 4 ($ps < .001$). All one-way ANOVAs for sessions 1-4 with group as independent variable yielded significant group differences. *Post hoc* (Bonferroni) analysis found that for session 1, Cantonese musicians significantly outperformed English musicians ($p < .001$) and English non-musicians ($p < .001$), but performed similarly as Cantonese non-musicians ($p = .139$). No significant differences were found for the other comparisons ($ps > .05$). For session 4, Cantonese musicians achieved significantly higher scores than Mandarin non-musicians, English musicians, and English non-musicians ($ps < .05$). Cantonese non-musicians significantly outperformed English musicians ($p < .05$) and English non-musicians ($p < .001$), but performed similarly as Mandarin musicians and Mandarin non-musicians ($ps = 1.000$). Mandarin musicians achieved significantly better scores than English musicians and English non-musicians ($ps < .01$), whereas Mandarin non-musicians also significantly outperformed English non-musicians ($p < .05$). No significant difference was found between English musicians and English non-musicians in all sessions ($ps = 1.000$).

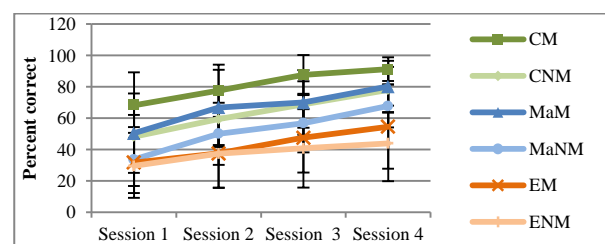


Figure 3: Mean percent correct scores by group in all session tests. Chance = 10%.

4. Discussion

4.1. Effect of musical experience

We found that all musicians outperformed non-musicians in the tone identification task, which involves lower linguistic level and no semantic processing. This suggests that musical experience facilitates tone identification regardless of L1 background.

In tone word learning task, which requires a higher linguistic level of processing, identifying, and mapping phonemic contrasts to lexical meaning at word level, English musicians performed similarly as Cantonese non-musicians in session 1 but performed significantly worse than Cantonese non-musicians in session 4. This demonstrated a reduction in musicality effect in later stages of tone word learning. This may provide support to previous studies that musicality effect has a dynamic role and its influence may vary depending on levels of contexts [13].

4.2. Effect of prosodic complexity

The effect of prosodic complexity was shown in both tone identification and tone word learning tasks. Specifically, for the tone identification task, Cantonese musicians performed better than English musicians, where no significant difference was found between Mandarin musicians and English musicians. In the tone word learning task, Cantonese musicians (non-musicians) outperformed English musicians, while Mandarin musicians (non-musicians) could only perform similarly as English non-musicians (musicians) in session 1 (session 4). These results suggest that tonal language background only appears to be advantageous when its prosodic complexity reaches a certain level, so that the existing perceptual system can be tuned to be similar to that non-native tonal system. Since Mandarin has the least tone system when compared to Thai and Cantonese, Mandarin speakers might not be as sensitive as Cantonese speakers to the specific features of Thai tones.

Furthermore, the positive effect of prosodic complexity can be extended to a higher linguistic level processing (i.e., tone word learning) and participants without musical background. Tonal perception and sound-to-meaning mapping ability are essential to tone word learning. The high prosodic complexity in Cantonese reduces loading in tonal perception, and allows Cantonese participants to allocate more attention to sound-to-meaning mapping. The prosodic complexity effect is thus amplified in tone word learning.

4.3. Interaction of musical and language experiences

We found no significant differences among Cantonese non-musicians, Mandarin non-musicians, and English non-musicians in tone word learning session 1, whereas Cantonese musicians could outperform English musicians. These results imply that musicianship may be necessary to activate the advantage of prosodic complexity effect in tone identification and initial stage of tone word learning. Also, we discovered that language experience modulates musicality effect on tone word learning based on the fact that Mandarin musicians could outperform Cantonese non-musicians, while English musicians could only attain similar performance as Mandarin non-musicians. Musicality effect was more robust in participants with tonal language background. This bidirectional relationship between musical and language experiences in tone identification provided further support to the overlap in musical and linguistic processing mechanisms as stated in previous research [18].

However, the most striking finding of our study is that the interaction between musical and language background varies in different contexts and stages of word learning. At the initial (session 1) and later (session 4) stages of tone word learning (see Figure 3), modulation of musicality effect by language background seemed not to persist as no significant differences

in performance between Mandarin musicians and Cantonese non-musicians, nor English musicians and Mandarin non-musicians were found. This suggests that musical experience facilitates word learning to a similar extent regardless of L1 background. Furthermore, it reveals that musical experience was no longer required to activate language effect in later stages of word learning such that Cantonese non-musicians and Mandarin non-musicians were also able to outperform English non-musicians. This is due to the increased importance of prosodic complexity effect and reduction in musicality effect towards higher level of contexts and later stages of word learning. In addition, taking tone word learning performances of all sessions into consideration, both Cantonese and Mandarin musicians and non-musicians outperformed English non-musicians. This suggests that musical experience is more important at initial stage of word learning and determines learning speed or initial success, whereas language background plays an important role in determining final/ later stages of success in word learning.

Our results also suggest that the effect of prosodic complexity of L1 seems to be more consistently advantageous, regardless of the extent, across different stages of learning. One possible explanation is that the fundamental and essential abilities involved in tone word learning are not simply general auditory processing and phonemic contrast identification, but also situating the contrasts at word level and mapping sound to semantic meaning. Although long-term musical experience enhances pitch acuity, the most essential factor remains whether there is an existing native tone system so that listeners can tune their native tone system, instead of developing an entirely new tonal inventory. Therefore, language background appeared to exert a more consistent beneficial impact on tone word learning than musical experience.

5. Conclusions

The current study extends previous studies by exploring the concept of L1 prosodic complexity, musical experience, and their interaction at both lower (tone identification) and higher-level (tone word learning) contexts.

We found that both musical experience and prosodic complexity play a role in non-native tone identification and tone word learning. Moreover, musical and language experience can indeed modulate each other, and such modulation exists in a bidirectional relationship. This interaction may also be modulated by the levels of processing involved in the context and stages of learning. Although the relative importance of these two effects varies across different contexts and stages of learning, prosodic complexity appears to be more consistently beneficial. These results illustrate that long-term musical experience can be transferred to a linguistic domain, where a more general auditory processing mechanism is adopted in lower-level linguistic context. Besides, native linguistic experience can be transferred to and be facilitative throughout different stages of non-native tonal language learning. Hence, the current results support an overlapping in musical-pitch and linguistic-pitch processing mechanisms, whereas the interaction of musical and language experience is far more complicated than expected. To conclude, this study corroborates with the study by Cooper and Wang [13] that both general auditory processing and linguistic knowledge are employed in non-native tone word learning, which points to an integrated tonal processing model being adopted.

6. References

- [1] J. Chobert and M. Besson, "Musical expertise and second language learning," *Brain Science*, vol. 3, pp. 923-940, 2013.
- [2] J. E. Flege, M. J. Munro, and I. R. A MacKay, "Factors affecting degree of perceived foreign accent in a second language," *Journal of Acoustical Society of America*, vol. 97, pp. 3125-3134, 1995.
- [3] L. R. Slevc and A. Miyake, "Individual differences in second language proficiency: does musical ability matter?" *Psychological Science*, vol. 17, pp.675-681, 2006.
- [4] P. Q. Pfordresher and S. Brown, "Enhance production and perception musical pitch in tone languages speakers," *Attention, Perception, & Psychologists*, vol. 71, no. 6, pp. 1385-1398, 2009.
- [5] Y. Lee, D. A. Vakoch, and L. H. Wurm, "Tone perception in Cantonese and Mandarin: A cross-linguistic comparison," *Journal of Psycholinguistic Research*, vol. 25, pp. 527-542, 1996.
- [6] R. P. Wayland and S. G. Guion, "Training English and Chinese listeners to perceive Thai tones: a preliminary report," *Language Learning*, vol. 54, pp. 681-712, 2004.
- [7] J. G. Barry and P. J. Blamey, "The acoustic analysis of tone differentiation as a means for assessing tone production in speakers of Cantonese," *Journal of the Acoustical Society of America*, vol. 116, pp. 1739-1748, 2004.
- [8] Y. R. Chao, *Language and Symbolic Systems*, Oxford, UK, Oxford University Press, 1948.
- [9] R. S. Bauer and P. K. Benedict, *Modern Cantonese Phonology*, Berlin, Germany, Mouton de Gruyter, 1997.
- [10] T. Hudak, *The World's Major Languages*, 2nd ed., London, UK, Routledge, 2009, pp. 660-676.
- [11] J. Alexander, P. C. Wong, and A. R. Bradlow, "Lexical tone perception in musicians and non-musicians," Interspeech 2005-Eurospeech: 9th European Conference on Speech Communication and Technology, Lisbon, Portugal, 2005, pp. 297-400.
- [12] T. L. Gottfried, "Music and language learning: effects of musical training on learning L2 speech contrasts," in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, Amerstadam, The Netherlands, John Benjamins, 2007, pp. 221-237.
- [13] A. Cooper and Y. Wang, "The influence of linguistic and musical experience on Cantonese word learning," *Journal of Acoustical Society of America*, vol. 131, no. 6, pp. 4756-4769, 2012.
- [14] P. C. Wong and T. K. Perrachione, "Learning pitch patterns in lexical identification by native English-speaking adults," *Applied Psycholinguistics*, vol. 28, pp. 565-585, 2007.
- [15] R. Wayland, E. Herrera, and E. Kaan, "Effects of musical experience and training on pitch contour perception," *Journal of Phonetics*, vol. 38, pp. 654-662, 2010.
- [16] N. Ellis and A. Beaton, "Factors affecting the learning of foreign language vocabulary: imagery keyword mediators and phonological short-term memory," *Quarterly Journal of Experimental Psychology*, vol. 46A, pp. 522-558, 1993.
- [17] A. Verma and M. Brysbaert, "A validated set of tool pictures with matched objects and non-objects for laterality research," *Literality: Asymmetries of Body, Brain and Cognition*, vol. 20, no. 1, pp. 22-48, 2015.
- [18] A. D. Patel, "Language, music, syntax, and the brain," *Nature Neuroscience*, vol. 6, pp. 674-681, 2003.