The role of musical experience in Cantonese lexical tone perception by native speakers of Thai

Angela Cooper and Yue Wang

Department of Linguistics, Simon Fraser University, Burnaby, B.C., Canada
akcooper@sfu.ca

Abstract

Adult non-native perception is subject to influence from a variety of factors, including native language and musical experience. The present study investigates the influence of these two factors in the perception and learning of non-native lexical tones. Native Thai-speaking musicians and non-musicians completed pre- and post-test identification tasks on five Cantonese tones, with 4 days of lexical identification training. Higher identification accuracy scores for musicians suggest that extensive experience with musical pitch enhances perception of non-native linguistic pitch. However, patterns of tonal accuracy improvement were similar across groups and can be attributed to the influence of the L1 tonal system.

Index Terms: lexical tone, musical experience, Cantonese, non-native perception

1. Introduction

1.1. Background

When infants are acquiring their native language (L1), they are open and receptive to a host of phonetic contrasts from other languages. After extensive L1 exposure, this perceptual receptivity declines to the point that it becomes significantly more challenging for adult learners to perceive these foreign contrasts [1]. However, learner performance can vary as a function of experience. Learners' L1 systems can interact with the developing second language (L2) phonetic system, shaping perception and the formation of new phonetic categories (e.g. [2]). Musical background can also potentially affect how efficiently and effectively non-native contrasts are acquired, as language and music may share common processing mechanisms [3-5].

Research has indicated that listeners' native phonetic systems can strongly influence the perception of novel sounds, and that the interaction of new phonetic structures, both segmental and suprasegmental, with established ones can have substantive effects on learning [6-8]. Concerning linguistic pitch, a study examining native English listeners' perception of Mandarin lexical tones suggested that the English intonational system had a direct influence on how listeners were perceiving particular tones, namely the Mandarin falling tone [6]. The authors argue that English declarative intonation and the Mandarin falling tone share acoustic similarities, thereby facilitating the identification of this tone (in isolation) as well as when it occurred in final position of a string of syllables. Similarly, Mandarin listeners' identification accuracy was best on the three Cantonese lexical tones that have similar counterparts in Mandarin, suggesting that native category representations can have a significant influence on non-native perception [7].

Additionally, it has been reported that perceptual discrepancies in lexical tone perception can be attributed to linguistic experience and language-specific weightings of two perceptual dimensions: F0 height and direction of change [8].

Results of the perception of Thai tones indicated that 'direction of change' appeared to be the most perceptually salient dimension for Thai listeners, whereas non-tone language listeners (English) gave greater weight to the 'height' dimension. It was posited that tone language and non-tone language groups are separable based on their respective rankings of these perceptual dimensions. Additional studies have also suggested that experience with being attuned to linguistically meaningful F0 changes and distinctions in one's native language can be particularly advantageous when perceiving non-native lexical tones, with tone language speakers having higher identification accuracy of non-native tones than non-tone language speakers [2, 9].

However, L1 experience is not the only factor to influence the ease with which some phonemic contrasts are acquired, as previous studies have pointed to a link between language and music (e.g. [10-11]). Both language and music qualify as rule-based systems where fundamental units (e.g. notes and phonemes) are arranged into higher-level hierarchical structures. Thus, it is not surprising that studies have suggested that language and music may have shared cognitive faculties and that crucial language areas in the brain are recruited during the processing of music [10], which may account for why verbal memory was also found to be better in musicians than non-musicians [12-13]. This close connection between language and music has led to a growing body of research addressing the effect of musical training on language learning. Given that both music and tone languages employ significant pitch modulations, several studies have examined the influence of musical experience on the acquisition of non-native suprasegmentals [3-5]. Research on the non-native perception and production of Mandarin lexical tones reported that musically-trained participants, with non-tone language backgrounds, performed significantly more accurately on tone identification and discrimination tasks [3-5].

1.2. The current study

This study explores the role of experience in non-native lexical tone perception. Native Thai-speaking musicians and non-musicians were asked to identify five Cantonese tones (High-Level, High-Rising, Low-Falling, Low-Rising, Low-Level) before and after they underwent a multi-session training program. As some studies have suggested that having a tone language background facilitates the acquisition of non-native tones [2, 9], the present research investigates the interaction of L1 with musicianship, and whether the addition of musical experience will still be advantageous for perceptual accuracy. In other words, this study seeks to extend the previous findings regarding the influence of musical experience on non-native suprasegmental learning to include listeners with a tone language background (Thai) identifying Cantonese tones.

We hypothesize that previous research will be extensible to tone language listeners and that musicians will have higher identification accuracy scores overall than non-musicians. Moreover, we hypothesize that L1 influence will likely

manifest in their patterns of tonal accuracy, such that the Cantonese lexical tones with close analogues in the Thai tonal system, including the high and low-level tones, will see the highest identification accuracy and the most improvement after training.

2. Methods

2.1. Participants

Thirty-three native Thai-speaking participants were involved in this study. All participants considered Thai to be their first and dominant language and reported having no previous knowledge of Cantonese or any other lexical tone language. Furthermore, they possessed normal hearing and cognitive abilities. They were recruited from Chulalongkorn University and Silpakorn University in Bangkok, Thailand. The participants were divided into two groups based on their musical experience. Eighteen participants (10 male, 8 female; mean age: 22 years) were considered "non-musicians", such that they had less than three years of musical experience (0.4 years on average) and no experience within the last five years. The "musician" group was comprised of fifteen participants (10 female, 5 male; mean age: 21 years), who had undergone at least six years of continuous Western instrumental music training (6-17 years, with a mean of 9 years) and had a current ability to play an instrument.

2.2. Stimuli

2.2.1. Pre-/post-test

Five CV monosyllables (*waj, low, si, pej, fu*) were produced with five Cantonese tones, creating a total of 25 real-word stimuli. The phonemes were common to both Thai and Cantonese in order to maintain focus on the suprasegmental information. The initial consonants selected were both voiced and voiceless to provide a variety of consonantal contexts, and five different vowels were included to ensure generalizability. The five Cantonese tones utilized in this study include the high-level, high-rising, low-falling, low-rising and low-level tones. The mid level-tone was not included, as it may be easily confused for the high and low level tones, particularly in the absence of any contextual cues [3]. Two native Cantonese speakers (1 male, 1 female) were recorded in a sound-attenuated booth. The stimuli were sampled at 44.1 kHz.

2.2.2. Training

Four novel speakers (2 male, 2 female) not used in the pre-/post-tests produced three CV monosyllables (*tsou, kwaaj, wu*) with five Cantonese tones. These 15 Cantonese words (3 syllables x 5 tones) were associated with meanings (common concrete nouns), as represented by a picture presented on the screen. Because participants would be receiving lexical identification training (learning sound-meaning pairings), these particular syllables were selected because they do not contain any semantic content in Thai, so as to reduce lexical competition with existing words in a participant's lexicon.

2.3. Procedure

2.3.1. Pre-/post-test Identification Task

The participants first completed a familiarization task in order to acquaint themselves with the task procedures. It also familiarized listeners with the five Cantonese tones and indicated how to identify them. They heard each Cantonese tone pronounced in isolation and viewed an appropriate tone image on the screen (a visual representation of the contour/level tone). Following these five trials, the participants were asked to respond after each stimulus, identifying the tone they heard by pressing the number on the computer keyboard corresponding to the appropriate tone image. Feedback on the accuracy of their response as well as the correct answer were provided. This task used productions of */ji/* by the female talker from the main task. Three randomized repetitions produced a total of 15 trials, lasting approximately 2 minutes.

The main task was a forced-choice identification task, where the participants identified the tone of each syllable from a selection of five options (presented as tone images on the screen), similar to the familiarization section. However, they did not receive any feedback on their identification accuracy. This task was comprised of four blocks, each containing the full set of 25 stimuli (5 syllables x 5 tones), equaling 100 trials. For each block, half of the stimuli were from the female speaker, the other half from the male speaker, such that across blocks, participants would hear all of the productions from both speakers twice. These trials were randomized within each block. The stimuli were presented with an inter-stimulus-interval of 3 seconds. The task took approximately 8 minutes.

2.3.2. Training

Participants received 4 days of training, where they learned the set of 15 training words and their associated meanings. Training listeners on sound-meaning pairings was utilized to simulate a more "natural" learning paradigm. The following stimulus presentation and testing procedures were modeled after training provided in [14-15].

A training session included listening to 5 blocks (each containing three words), while viewing the visual representation of their meaning. Each block of words contained non-minimal triplets (i.e. three different tones on three different syllables). Every block concluded with a small quiz on the three words learned in that block, whereby participants heard a stimulus and were presented with the three pictures of the words they had just learned. They were asked to indicate the correct meaning for the word by selecting the appropriate picture. They were provided with feedback, informing them whether or not their answer was correct, displaying the correct answer and re-playing the stimulus. Participants received two additional blocks reviewing the training items. The first review was comprised of all 15 words produced by one female speaker from training, blocked by syllable. Blocking for syllable enabled participants to hear minimal tone quintuplets in succession in order to highlight the tonal distinctions. The second review consisted of all 15 words produced by 2 speakers (1 male, 1 female), where participants chose the meaning after each stimulus by selecting the appropriate picture from all 15 options on the screen. Similar to the block quizzes, both review blocks provided feedback after each response. At the end of each session, participants were tested on all 15 words learned in the training program without feedback. The session test followed the same format as the final review block, now involving all 15 training words produced by all 4 speakers. The total duration of a training session was approximately 20 minutes. Each participant received 2 training sessions per day, with approximately 15 minutes between each training session. Each day of training was separated by 2-4 days.

3. Results

The mean percent correct pre-test and post-test identification scores for the non-musicians and musicians are provided in

Figure 1. A 3-way mixed analysis of variance (ANOVA) was conducted, with Group (non-musician, musician) as a between-subjects factor, and Test (pre, post) and Tone (high-level, high-rising, low-falling, low-rising, low-level) as repeated measures.
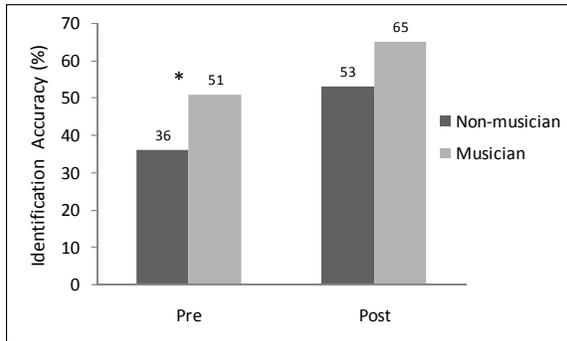

Figure 1: *Mean identification accuracy for pre- and post-tests by Thai musicians and non-musicians. "*": statistically significant difference (p<0.05).*

A significant main effect of Group was found, with the musicians obtaining higher identification accuracy scores than the non-musicians overall (58% vs. 45%) $[F(1,31)=7.94, p=.008]$. As expected, a significant main effect of Test was also obtained $[F(1,31)=45.60, p<.0001]$, with both musicians and non-musicians significantly increasing their tonal identification accuracy following training across tones (14% and 17% increases from pre- to post-test, respectively). Further analysis for each test with Group (musician, non-musician) as a between-subjects factor indicated that the musician group performed significantly better on the pre-test identification task than the non-musician group (51% vs. 36%) $[F(1,31)=13.57, p=.001]$, but that there was no longer any significant group difference on the post-test (65% vs. 53%) $[F(1,31)=3.66, p=.065]$ (cf. Figure 1).

The ANOVA also yielded a significant main effect of Tone $[F(4,31)=13.78, p<.0001]$. The musician and non-musician groups were largely consistent with which tones they found challenging and which easier to identify. Post-hoc analysis (Bonferroni adjusted) revealed that both groups found the low-rising $(p<.0001)$ and high-rising $(p<.0001)$ tones to be significantly more challenging than the low-falling and high and low-level tones.

Additionally, a significant 3-way interaction of Group x Test x Tone was also observed $[F(4,31)=2.71, p=.033]$. Two-way mixed ANOVAs for each tone were performed, with Test (pre, post) as repeated measures, and Group (musician, non-musician) as a between-subjects factor. Figure 2 illustrates mean percent correct by Group and Test for each tone. Significant main effects for Group were found for the high level $(p=.007)$, high $(p=.018)$ and low rising tones $(p=.007)$, indicating that the musicians identified all of these tones more accurately than the non-musicians before and after training. A Group x Test interaction was found for the low-level tone $(p=.009)$. Further analysis revealed that musicians were significantly more accurate for this tone on the pre-test, but no significant group differences were found on the post-test. There were no significant group differences for the low-falling tone across tests.

Furthermore, a 2-way ANOVA with Group as a between-subjects factor and Tone as repeated measures was performed with percent improvement as the dependent variable. As illustrated by Figure 3, no significant group differences in the amount of improvement from pre- to post-test were found across tones. The only significant group difference was in the low-level tone $[F(1,31)=7.738, p=.009]$, in that the non-musicians significantly increased in accuracy, but no significant increase was found for the musicians. Across groups, the low-falling and high-level tones saw the most improvement from pre-test to post-test. Additionally, the high-rising and low-rising tones saw the least amount of improvement.
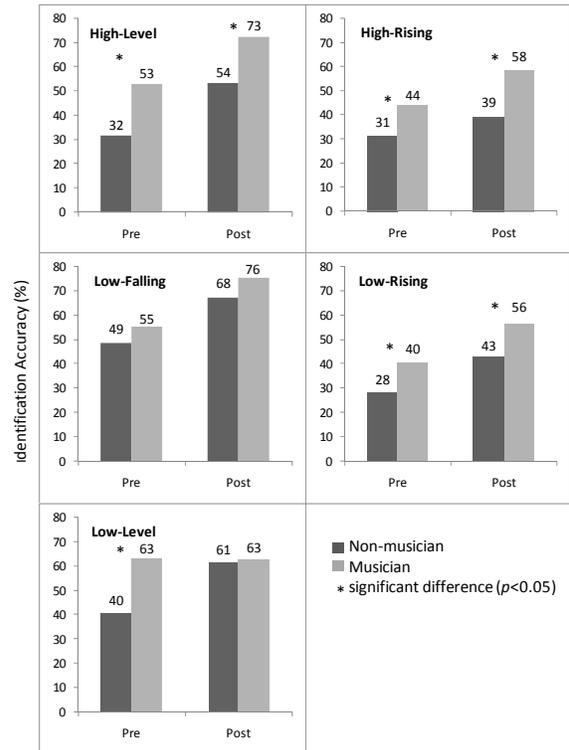

Figure 2: *Mean identification accuracy by tone for pre- and post-tests by musicians and non-musicians*
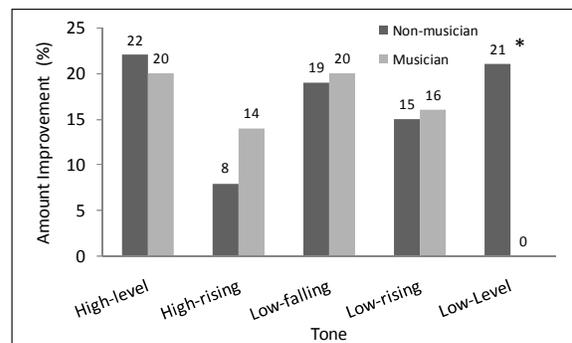

Figure 3: *Mean percentage of tone improvement from pre-test to post-test for musicians and non-musicians. "*": statistically significant difference (p<0.05).*

4. Discussion

With regards to group differences, it appears that having a tone language background does not render the influence of musical experience inconsequential, extending previous findings that musical pitch processing experience can enhance linguistic

pitch acuity to include musicians with a tonal L1. The results demonstrate that Thai musicians were significantly more accurate at identifying non-native lexical tones than Thai non-musicians overall, which is consistent with previous findings on the influence of musical experience of non-native suprasegmental perception [3-5]. Musicians performed significantly better on 4 out of the 5 tones on the pre-test; however, there were no significant group differences on the post-test for the low-level and low-falling tones. In this case, the musicians may have started to reach a ceiling, with respect to the amount of fine-grained phonetic information they could receive from this particular type of training.

It appears that musicianship did not provide a significant advantage in terms of the amount of improvement, as both groups saw similar levels of improvement after training. This may be due to the type of training, in that its primary focus was not tuning the listeners into the fine-grained acoustic pitch distinctions, but providing more holistic training. The goal was to situate training in a more linguistically-meaningful context, where listeners would need to learn the tonal, segmental and semantic distinctions between the words. Despite training not being tone-centered, both musician and non-musician groups saw significant improvements in tone identification accuracy from pre-test to post-test.

A similar pattern of tonal accuracy and improvement across musician and non-musician groups points to an influence of the L1 tonal system on non-native perception. The results of the tonal analysis revealed that the high-level tone saw the most improvement for both groups. This could be explained by the fact that it is acoustically similar, in both height and contour shape, to the high-level Thai tone. The low-falling tone saw a mean 20% improvement, which could be attributed to listeners detecting the falling contour direction and identifying it with their L1 falling tone. This analysis would be consistent with findings that non-native tones acoustically similar to Mandarin tones saw the most improvement for Mandarin listeners [7]. Given the presence of a rising tone in their native language, one would expect the high-rising tone to see substantive improvements for Thai listeners. However, the relatively small improvement in accuracy after training is likely due to the presence of the second rising tone (low-rising). This challenge for listeners may stem from the fact that one native rising tone category is forced to map onto two rising-tone categories in Cantonese. Such findings are consistent with previous research positing that two possible exemplars from a single native phoneme category will be particularly difficult to discriminate (e.g. [16]). While these findings cannot be conclusively linked to the influence of the L1 tonal system without additional language groups or an L1-L2 mapping test, a follow-up study is underway with English-speaking participants. The addition of another language group should allow us to compare their patterns of tonal accuracy to determine if they are a result of L1 influence.

5. Conclusions

The present research investigated the role of musical experience in listeners with a tonal L1 background on the perception of non-native lexical tones. The results showed that Thai-speaking musicians were more accurate at identifying Cantonese tones than non-musicians. This provides further support for a growing body of research on the transferability of musical experience into the linguistic domain, in that experience with musical pitch modulations can extend into proficiency with linguistic pitch. However, both groups displayed similar patterns of tonal accuracy improvement, with respect to which tones they found easiest or most challenging, suggesting that L1 tonal categories influence the ease with which non-native tonal representations are constructed. Our ongoing research investigates the integral effects of L1 and musical experience by including an additional group of native English speakers. This will allow us to tease apart the relative influences of these factors and provide a more complete picture of the role of L1 in non-native perception.

7. References

[1] Werker, J. and Tees, R. C., "Cross-language speech perception: Evidence for perceptual reorganization during the first year of life", Infant Behavior & Development, 25:121-133, 2002.

[2] Lee, Y-S., Vakoch, D. A. and Wurm, L. H., "Tone Perception in Cantonese and Mandarin: A Cross-Linguistic Comparison", Journal of Psycholinguistic Research, 25:527-542, 1996.

[3] Alexander, J. A., Wong, P. C., and Bradlow, A., "Lexical Tone Perception in Musicians and Non-musicians", in Proceedings of Interspeech 2005, Lisbon, Portugal, 2005.

[4] Delogu, F., Lampis, G., and Belardinelli, M. O., "Music-to-language transfer effect: may melodic ability improve learning of tonal languages by native nontonal speakers?", Cognitive Process, 7:203-207.

[5] Gottfried, T. L., "Music and language learning: effects of musical training on learning L2 speech contrasts", in O. Bohn & M. J. Munro [Eds.], Language Experience in Second Language Speech Learning: In honor of James Emil Flege, 221-237, Philadelphia: John Benjamins Publishing Company, 2007.

[6] Broselow, E., Hurtig, R. and Ringen, C., "The perception of second language prosody", in S. Weinberger & G. Ioup [Eds], Interlanguage Phonology: The Acquisition of a Second Language Sound System, 350-361, Cambridge: Newbury House Publishers, 1987.

[7] Francis, A., Ciocca, V., Ma, L. and Fenn, K., "Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers", Journal of Phonetics, 36:268-294, 2008.

[8] Gandour, J., "Tone perception in Far Eastern languages", Journal of Phonetics, 11:149-175, 1983.

[9] Wayland, R. P. and Guion, S. G., "Training English and Chinese Listeners to Perceive Thai Tones: A Preliminary Report", Language Learning, 54:681-712, 2004.

[10] Besson, M. and Schön, D., "Comparison between language and music", Annals of the New York Academy of Sciences, 930, 2001.

[11] Patel, A.D. Music, Language, and the Brain. Oxford: Oxford University Press, 2008.

[12] Chan, A. S., Ho, Y. and Cheung, M., "Music training improves verbal memory", Nature, 396:128, 1998.

[13] Brandler, S. and Rammsayer, T. H., "Differences in Mental Abilities between Musicians and Non-Musicians", Psychology of Music, 31(2):123-128, 2003.

[14] Curtin, S., Goad, H. and Pater, J. V., "Phonological transfer and levels of representation: the perceptual acquisition of Thai voice and aspiration by English and French speakers", Second Language Research, 14:389-405, 1998.

[15] Wong, P.C. and Perrachione, T. K., "Learning pitch patterns in lexical identification by native English-speaking adults", Applied Psycholinguistics, 28:565-585, 2007.

[16] Best, C. T., McRoberts, G. W. and Goodell, E., "Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system", Journal of the Acoustical Society of America, 109:775-794, 2001.