The Selectivity of Musical Advantage:

Musicians Exhibit Perceptual Advantage for Some but not All Cantonese Tones

William Choi

The Education University of Hong Kong, Taipo, Hong Kong

Abstract

The OPERA hypothesis theorizes how musical experience heightens perceptual acuity to lexical tones. One missing element in the hypothesis is whether musical advantage is general to all or specific to some lexical tones. To further extend the hypothesis, this study investigated whether English musicians consistently outperformed English nonmusicians in perceiving a variety of Cantonese tones. In an AXB discrimination task, the musicians exhibited superior discriminatory performance over the nonmusicians only in the high level, high rising, and mid-level tone contexts. Similarly, in a Cantonese tone sequence recall task, the musicians significantly outperformed the nonmusicians only in the contour tone context but not in the level tone context. Collectively, the results reflect the selectivity of musical advantage—musical experience is only advantageous to the perception of some but not all Cantonese tones, and elements of selectivity can be introduced to the OPERA hypothesis. Methodologically, the findings highlight the need to include a wide variety of lexical tone contrasts when studying music-to-language transfer.

The Selectivity of Musical Advantage:

Musicians Exhibit Perceptual Advantage for Some but not All Cantonese Tones

Music training induces long-term psychological and cognitive benefits (Ho, Cheung, & Chan, 2003; Paquette & Goulet, 2014; Rodrigues, Loureiro, & Caramelli, 2010; Schellenberg & Mankarious, 2012; Stoesz, Jakobson, Kilgour, & Lewycky, 2007). In the language domain, one eminent advantage of music training is enhanced perceptual sensitivity to fundamental frequency-based phonological features (*lexical tones*, Burnham & Mattock, 2007; Gandour, 1981) (e.g., Alexander, Wong, & Bradlow, 2005; Delogu, Lampis, & Olivetti-Belardinelli, 2006; Kraus & Chandrasekaran, 2010; Patel, 2011). Given the limited number of lexical tones tested in previous studies, a question naturally arises as to whether musical advantage is general to all or specific to some lexical tones. To fill this important research gap and advance the theoretical understanding of music-to-language transfer, this study investigated whether musicians consistently exhibited perceptual advantage over nonmusicians across all Cantonese tone contexts.

The OPERA hypothesis posits that music training enhances the neural encoding of speech as long as five conditions (overlap, precision, emotion, repetition, and attention) are met (Patel, 2011). For example, in terms of precision, Patel argues that music must entail more fine-grained processing than speech for music-to-language transfer to occur. From the theoretical perspective, the OPERA hypothesis has offered a comprehensive account for a large body of studies that demonstrated superior lexical tone perception in musicians vis-à-vis nonmusicians. In an early study, Italian listeners who had high melodic ability were better at detecting tonal variations than those who had low melodic ability (Delogu et al., 2006). Musical advantage was also evident in AX and identification tasks—English-speaking musicians were more accurate than English-speaking nonmusicians in identifying and

discriminating Mandarin tones (Alexander et al., 2005). Remarkably, the English-speaking

musicians even performed on par with native listeners in Mandarin tone discrimination. This

reflected that music training could boost non-native listeners' tonal sensitivity to a native

level. English-speaking musicians' advantage in Mandarin tone perception was even

identified at the phrase level (Zheng & Samuel, 2018). The central idea from the studies was

that music training could enhance listeners' ability to discriminate lexical tones in AX tasks,

at least for speakers of nontonal languages (e.g., English and Italian).

Musical advantage was also identified at higher perceptual levels. In a Cantonese

tone-word learning study, listeners were trained over seven sessions to identify words

differing minimally by five Cantonese tones (Cooper & Wang, 2012). In the seventh training

session, the English-speaking musicians identified tone-words more accurately than the

English-speaking nonmusicians, even though both groups performed similarly in the first

session. The results reflected that music training was not only beneficial to low level speech

discrimination. Crucially, music training also brought about positive impacts in tasks that

entailed higher level perceptual operations such as memory encoding and storage of lexical

tones. This musical advantage emerged upon adequate tone-word learning experience.

Although musical advantage has been well studied, one important element is still

missing in the literature: is musical advantage general to all or specific to some lexical tones?

As described previously, the OPERA hypothesis has put forward five prerequisites for music-

to-language transfer, but there is a lack of explicit specification of whether the perceptual

benefits resulting from music-to-language transfer is "selective" to lexical tones (Patel,

2011). Apart from the lack of relevant analysis conducted, the available data in the literature

could not adequately address this question given the small number of tones tested. For

example, a previous study only tested the discrimination of level tones but not contour tones

(Zhang & Samuel, 2018). Even in the studies that utilized all possible lexical tone pairs (e.g.,

Alexander et al., 2005), the number of tone contrasts was still limited given the small number of lexical tones in Mandarin (Gu & Lee, 2009).

With a rich tonal repertoire, Cantonese provides a fascinating window into the potential interaction between musical experience and lexical tones in lexical tone perception. According to Chao (1930), there are six lexical tones in Cantonese, namely high level (HL, 55), high rising (HR, 25), mid-level (ML, 33), low falling (LF, 21), low rising (LR, 23), and low level tones (LL, 22).[1] Acoustically, HL, HR, and ML tones have higher average F0 (i.e, pitch height) than LL, LF. and LR tones (see Figure 1). In terms of the changes of overall F0 over time (i.e., F0 contour), HR, LF, and LR tones exhibit more drastic changes while HL, ML, and LL tones remain relatively steady. Linguistically, variations of lexical tones in a Cantonese syllable can result in multiple meanings, e.g., 優 /jau-HL/ *distinction*, 柚 /jau-HR/ *grapefruit*, 幼 /jau-ML/ *thin*, 油 /jau-LF/ *oil*, 有 /jau-LR/ *have*, 右 /jau-LL/ *right*. The rich tonal repertoire gives rise to a much larger set of lexical tone contrasts in Cantonese (15) than in Mandarin (6) (Choi, Tong, Gu, Tong, & Wong, 2017; Choi, Tong, & Singh, 2017; Gandour, 1981).

Insert Figure 1 about here

In this study, it is hypothesized that the musical advantage in Cantonese tone perception is selective. Drawing on parallels with congenital amusia, there is available evidence suggesting that amusical disadvantage in lexical tone perception is not purely general but instead dependent on specific lexical tone contrasts (e.g., Tillmann, Burnham, Nguyen, Grimault, Gosselin, & Peretz, 2011). In Tillmann and colleagues' study, French congenital amusics were less accurate than French controls in discriminating all Thai tone contrasts. This seemed to indicate a blanket disadvantage in Thai tone discrimination.

However, a fine-grained analysis further revealed a significant interaction between group and tone contrasts. Although the French congenital amusics performed consistently poorer than the French controls, the performance gap narrowed for the rising-and-falling tone contrast. This interaction suggested that the perceptual disadvantage was dependent on specific tone contrasts. A similar pattern was found among Mandarin congenital amusics; they identified Mandarin tones consistently poorer than the controls but the performance gap was wider for mid-rising tone (Nan, Sun, & Peretz, 2010). Collectively, studies on congenital amusics indicated that amusia exerted differential effects on different lexical tones, even though the congenital amusics performed poorer than the controls as a whole.

Further indirect evidence for the hypothesis could be drawn from cross-linguistic research. While some early studies reported that tonal listeners were better at non-native tone discrimination than nontonal listeners (e.g., Lee, Vakoch, & Wurm, 1996; Wayland & Guion, 2004), a recent study with a more nuanced analysis demonstrated a clear specificity: Thai listeners outperformed English listeners only in identifying two out of four Mandarin tones (Li, 2016). Paralleling the cases of congenital amusics (Nan et al., 2010; Tillmann et al., 2011), the interaction between group and lexical tone suggested that tone-language advantage was selective rather than general. Taken together, research on congenital amusia and cross-linguistic transfer indicated that the perceptual (dis)advantages associated with amusia/language experience were hardly general. Thus, it is reasonable to hypothesize that the musical advantage in lexical tone perception is selective.

To further extend the OPERA hypothesis and advance the theoretical understanding of music-to-language transfer, this study investigated whether the musical advantage in Cantonese tone perception was general or selective. Of particular interest to this study was whether English musicians consistently outperformed English nonmusicians across different Cantonese tone contexts. Given that musical advantage was evident in both low level

discrimination and high level perceptual operations (Alexander et al., 2005; Cooper & Wang, 2012; Delogu et al., 2006; Zheng & Samuel, 2018), listeners were tested with both AXB discrimination and sequence recall tasks.

Method

Participants

Forty English listeners were recruited from University College London and its surrounding areas. The participants were divided into two groups (i.e., musicians and nonmusicians), each with 20 participants. Adopting the same criteria from a previous study, the musicians were those who had received at least seven years of continuous music training and had the ability to play their instruments at the time of testing (Tong, Choi, & Man, 2018). The nonmusicians were those who (a) had received less than two years of music training, if any, throughout their life, (b) had not received any music training in the past five years and (c) could not play any musical instrument at the time of testing. One nonmusician did not show up for the experiment, and one nonmusician was excluded from the study for having received five years of music training. One musician was excluded from the study due to self-reported Mandarin learning experience. The final sample consisted of 19 musicians (5 males, 14 females) and 18 nonmusicians (8 males, 10 females).

The mean ages of the musicians and nonmusicians were 26.63 years ($SD = 5.89$ years) and 32.67 years ($SD = 11.60$ years) respectively. The mean onset age of music training was 7.84 years ($SD = 2.89$ years) for musicians and 12.00 years ($SD = 4.86$ years) for nonmusicians. On average, the musicians had received 11.63 years of music training ($SD = 3.90$ years) while the nonmusicians had only received 0.90 year of music training ($SD = 1.56$ year). According to self-reports, none of the participants possessed absolute pitch.

Tasks

Cantonese tone discrimination task. Adopting an AXB paradigm, this task assessed the ability to discriminate Cantonese tones. The stimuli included all possible 15 pairs of Cantonese tone contrasts[2] (i.e., HL-HR, HL-ML, HL-LF, HL-LR, HL-LL, HR-ML, HR-LF, HR-LR, HR-LL, ML-LF, ML-LR, ML-LL, LF-LR, LF-LL, LR-LL, embedded in different segments, e.g., /ka/, /ku/, /kɛ/, /ki/ and /kɔ/). All stimuli were naturally recorded by two native Cantonese speakers, one male and one female. On each trial, three syllables (e.g., /ka-HL/ /ka-ML/ /ka-ML/) were presented via Sennheiser HD280 PRO headphones, with an interstimulus interval of 600 ms. Participants then indicated, by pressing the associated keys, whether the first or the third syllable carried the same lexical tone as the second syllable did. To prevent the participants from relying on simple acoustic comparisons, AB and X were produced by speakers of different genders. There were eight trials for each Cantonese tone contrast and the total number of trials was 120. The task began with six practice trials with feedback provided. The sample-specific internal consistency was high ($\alpha = .81$).

Sequence recall task. This task, which assessed the ability to represent and store Cantonese tones in memory, was modified from a stress sequence recall task (Dupoux, Peperkamp, & Sabastian-Galles, 2010). There were three contexts—a vowel context (/ta-HL/ /tɛ-HL/), a level tone context (/ta-HL/-/ta-ML/), and a contour tone context (/ta-LF/-/ta-LR/). All stimuli were recorded by the same native Cantonese speakers as above. In each context, the first nonword (e.g., /ta-LF/) was associated with the key [1] while the second nonword (e.g., /ta-LR/) was associated with the key [2].

In the familiarization phase, the participants listened to the items as many times as desired by pressing the keys [1] and [2]. Following the familiarization phase, there were eight identification trials of [1] and [2] with feedback provided.

On each trial in the contour tone context, a sequence of syllables (e.g., /ta-LR/ /ta-LF/ /ta-LF/ /ta-LR/) with a varying length (2, 3, 4, 5, or 6) was presented. Following sequence presentation, participants reproduced the sequence by pressing the associated keys in the correct order (e.g., 2112 for /ta-LR/ /ta-LF/ /ta-LF/ /ta-LR/). They could also skip any trial by simply pressing enter. A response was marked as correct only if it was a 100% match of the sequence presented. Any other responses were marked as incorrect. One point was given for each correctly reproduced sequence. There were six trials for each sequence length, giving rise to 30 trials in the contour tone context. The same procedure was adopted in the vowel and level tone contexts. The sample-specific internal consistencies were high in the vowel context ($\alpha = .84$), level tone context ($\alpha = .89$) and contour tone context ($\alpha = .94$).

Short-term memory task. Adopted from previous studies (Choi, Tong, & Samuel, 2019; Zheng & Samuel, 2018), this task assessed nonverbal short-term memory. On each trial, a sequence of colors (e.g., red-green-blue) was presented in an object containing four wedges (red, green, blue, and yellow). Participants then reproduced the color sequence by clicking on the corresponding wedges. One point was given for each correctly reproduced sequence. The sequence length increased by one following each correct response. The score would increase by one until the participant failed to produce the sequence. Participants completed the task for five times and the median score was obtained. The sample-specific internal consistency was satisfactory ($\alpha = .76$).

Nonverbal intelligence task. Used in previous studies (Choi et al., 2019; Zheng & Samuel, 2018), this task provided a quick estimate of participants' nonverbal intelligence. There were 14 multiple choice questions chosen from two intelligence tasks online (http://www.iq-test.com/free-iq-test/ and http://www.quickiqtest.net). For each question, a picture with an incomplete visual pattern was presented. It was followed by multiple pictures,

one of which could complete the visual pattern. One point was given for each correct answer.

The sample-specific internal consistency was satisfactory (α = .71).

<div align="center">Results</div>

Control Variables

To examine whether the musicians and the nonmusicians differed in any of the

control variables, a multivariate analysis of variance (MANOVA) was conducted with

nonverbal intelligence and short-term memory, $r(35)$ = .46, $p < .01$, being the dependent

variables and group being the independent variable. The overall group effect was not

significant, $\Lambda_{\text{Wilks'}}$ = .97, $F(2, 34)$ = 0.53, p = .594, indicating that the two groups matched on

nonverbal intelligence and short-term memory. Thus, the two control variables were not

included in the subsequent analyses.

Cantonese Tone Discrimination Task

Behavioral analysis. The main research question was whether musical advantage was

selective to lexical tones. As shown in Figure 2, the musicians appeared to have superior

discriminatory performance than the nonmusicians in the HL, HR, and ML tone contexts, but

the performance gaps were narrow in the LF, LR, and LL tone contexts. To statistically

address the research question, a two-way mixed analysis of variance (ANOVA) was

conducted on mean accuracy with tone context (HL, HR, ML, LF, LR, and LL) being the

within-subject factor and group (musician and nonmusician) being the between-subjects

factor. The analysis yielded a significant main effect of tone context, $F(5, 175)$ = 16.22, $p <$

.001, η_p^2 = .32, and a marginally significant main effect of group, $F(1, 35)$ = 3.67, p = .064,

η_p^2 = .10, reflecting a potential musical advantage in lexical tone discrimination. Consistent

with the hypothesis, there was a significant interaction between tone context and group, $F(5,$

$175)$ = 3.00, $p < .05$, η_p^2 = .08, indicating that the musicians significantly outperformed the

nonmusicians only in discriminating some but not all Cantonese tones. Pairwise comparisons further revealed that the musicians significantly outperformed the nonmusicians in the HR tone context, $p < .05$. Additionally, the musicians marginally outperformed the nonmusicians in the HL tone context, $p = .085$, and the ML tone context, $p = .082$. Both groups performed similarly in the LF tone context, $p = .197$, the LR tone context, $p = .648$, and the LL tone context, $p = .260$.

Insert Figure 2 about here

Acoustic analysis. All produced tokens were analyzed acoustically with Praat 5.4.02 (Institute of Phonetic Sciences, University of Amsterdam, The Netherlands). The individual and token-averaged duration, F0 height, minimum F0, maximum F0, F0 onset, F0 offset, and F0 contour are summarized in Tables 1-3. To test whether the acoustic features differed significantly as a function of lexical tones, a two-way mixed ANOVA was conducted with acoustic parameter being the within-subject factor and tone being the between-subjects factor. There were significant main effects of tone, $F(5, 54) = 11.58$, $p < .001$, $\eta_p^2 = .52$, and acoustic parameter, $F(6, 324) = 767.62$, $p < .001$, $\eta_p^2 = .93$. Most importantly, the interaction between tone and acoustic parameter was significant, $F(30, 324) = 3.37$, $p < .001$, $\eta_p^2 = .24$. This indicated that the acoustic features, when averaged across tokens, differed across Cantonese tones (see Figures 3 and 4). For example, HL tone had a higher F0 height than LF, LL (*ps* < .01), HR ($p = .071$), and LR tones ($p = .058$). The remaining results are summarized in Appendix A (see Supplementary Materials).

Insert Tables 1, 2 and 3 about here

Insert Figures 3 and 4 about here

The second goal of this study was to identify the potential acoustic underpinnings of the selectivity of musical advantage. For each instance of a tone contrast, the acoustic

differences between the contrastive tones were obtained, yielding eight tokens for each tone

contrast and 40 tokens in each lexical tone context. Correlational analyses were conducted to

identify the potential associations between the acoustic differences and the mean accuracies

of (i) the musicians and (ii) the nonmusicians (see Table 4).

Insert Table 4 about here

Of interest to this study, in the HR tone context where musical advantage was found,

correlational analyses yielded marginally significant correlations between F0 contour and the

mean accuracy of the musicians, $r(38) = .28$, $p = .083$, and between F0 onset and the mean

accuracy of the nonmusicians, $r(38) = .29$, $p = .072$. These imply that the musicians and the

nonmusicians attended to different acoustic cues in the HR tone context. Similarly, in the HL

tone context, correlational analyses revealed significant associations between the mean

accuracy of the musicians and duration, $r(38) = .40$, $p < .05$; F0 height, $r(38) = .42$, $p < .01$;

maximum F0, $r(38) = .35$, $p < .05$; F0 onset, $r(38) = .43$, $p < .01$; and F0 offset, $r(38) = .33$, p

$< .05$, and marginally significant correlations between the mean accuracy of musicians and

minimum F0, $r(38) = .28$, $p = .081$; and F0 contour, $r(38) = .30$, $p = .056$. Contrastively, only

a marginally significant correlation was found between the mean accuracy of the

nonmusicians and F0 contour, $r(38) = .27$, $p = .096$, implying that the nonmusicians attended

to F0 contour in the HL tone context. In the ML tone context, a marginally significant

association was found between the mean accuracy of the musicians and F0 offset, $r(38) = .29$,

$p = .068$, and between the mean accuracy of the nonmusicians and F0 contour, $r(38) = .28$, p

$= .076$, also suggesting that the two groups relied on different acoustic cues.

Interestingly, in the LR tone context where musical selectivity was absent, the two

groups attended to the same acoustic cues. In particular, correlational analyses yielded

significant associations between F0 contour and the mean accuracy of the musicians, $r(38) =$

.44, $p < .01$, and the mean accuracy of the nonmusicians, $r(38) = .37$, $p < .05$. F0 onset also

marginally correlated with the mean accuracy of the musicians, $r(38) = .29$, $p = .068$, and the

mean accuracy of the nonmusicians, $r(38) = .31$, $p = .052$. Lastly, F0 contour correlated

significantly with the mean accuracy of the musicians in the LF tone context, $r(38) = .40$, $p <$

.05, and the LL tone context, $r(38) = .40$, $p < .05$.

Sequence Recall Task

A further goal was to examine whether the selectivity of musical advantage was also

present at higher perceptual levels. As shown in Figure 5, the musicians appeared to

outperform the nonmusicians across all contexts, implying the universality of musical

advantage across different tone contrasts. To statistically test the hypothesis, a 2-way mixed

ANOVA was conducted on mean accuracy with context (vowel, level tone, and contour tone)

being the within-subject factor and group (musician and nonmusician) being the between-

subjects factor. The analysis revealed significant main effects of context, $F(2, 70) = 61.60$, p

$< .001$, $\eta_p^2 = .64$, and group, $F(1, 35) = 7.07$, $p < .05$, $\eta_p^2 = .17$, reflecting a musical

advantage. In line with the hypothesis, there was a significant interaction between context

and group, $F(2, 70) = 7.15$, $p < .01$, $\eta_p^2 = .17$. Further analysis showed that the musicians

outperformed the nonmusicians only in the contour tone context, $p < .01$, but not in the

vowel, $p = .180$, and level tone contexts, $p = .142$. The results suggest that musical advantage

is selective to tones even in a more complex perception task. The lack of musical advantage

in the vowel context also reflects that the above found musical advantage is not speech

general but specific to (contour) tones.

Insert Figure 5 about here

Discussion

This study endeavored to investigate whether music-to-language transfer was general to all or specific to some Cantonese tones. Consistent with the hypothesis, musical advantage was found only in some but not all Cantonese tone contexts. Remarkably, the selectivity of musical advantage was evident in both AXB discrimination and sequence recall tasks.

The most important finding in the current study is the selectivity of musical advantage. As reviewed in the Introduction, previous studies consistently reported that musicians outperformed nonmusicians in lexical tone perception, providing convergent evidence for the positive impact of musical experience on lexical tone perception (Alexander et al., 2005; Cooper & Wang, 2012; Delogu et al., 2006; Zheng & Samuel, 2018). The present study extends the previous studies by demonstrating that the effect of musical experience on lexical tone perception is not general but differential. Specifically, the English musicians exhibited perceptual advantages over the English nonmusicians only in half of the Cantonese tone contexts. From the theoretical perspective, the OPERA hypothesis argued that music training could enhance neural precision for speech encoding, but there was a lack of specification of whether music-to-language transfer was general or specific (Patel, 2011). The selectivity of musical advantage found herein suggests a need to expand the OPERA hypothesis by incorporating a mechanism that governs how musical experience interacts with phonemes/lexical tones in terms of their phonetic and acoustic properties. Drawing on parallels with the language domain, in the Perceptual Assimilation Model, whether first-language experience is beneficial or detrimental to non-native speech discrimination depends on how the non-native phonemes are assimilated into first-language phonemic categories (Best, 1995; Best & Tyler, 2007; Reid et al., 2015). The second step towards a new model or refined OPERA hypothesis is to explore how some specific phonetic/acoustic properties make certain lexical tones more relevant to musical experience than the other lexical tones do.

Although the specific mechanism governing the selectivity of musical advantage remains unclear, the current results can help to rule out some possible mechanisms. In the Cantonese tone discrimination task, although the musicians had better performance than the nonmusicians in the HL and ML tone contexts, no obvious musical advantage was found in the other level tone context (i.e., the LL tone context). Similarly, for the contour tones, despite the clear musical advantage in the HR tone context, the musicians and nonmusicians performed similarly in the LF and LR tone contexts. Based on the above results, the selectivity of musical advantage does not appear to be governed by the typology (level or contour) of the lexical tones. Apart from typology, the perceptual difficulty of the individual tone contrasts does not seem to drive the selectivity of musical advantage. Indeed, different tone contrasts have different levels of difficulty for English listeners, but no consistent pattern between musical advantage and perceptual difficulty has been observed in the present and previous studies (Qin & Mok, 2011).

That said, the acoustic-behavioral correlational analyses have provided some useful hints about the mechanism governing the selectivity of musical advantage. Despite the close acoustic resemblance between HR and LR tones (see Mok, Zuo, & Wong, 2013), musical experience only favored HR but not LR tone perception. Interestingly, the acoustic-behavioral analyses showed that the musicians attended to F0 contour (a more effective cue for HR tone) while the nonmusicians relied on F0 onset (a less effective cue for HR tone) for HR tone perception. Thus, it is possible that the selectivity of musical advantage is driven by the use of acoustic cues. In particular, it might be the case that musical experience had orientated the musicians to a more reliable acoustic cue (i.e., F0 contour), for HR tone perception. Contrastively, given that both musicians and nonmusicians relied on F0 contour for perceiving LR tone, musical experience had very little facilitative effect on LR tone

perception. This potential account for the selectivity of musical advantage is worthy of further examination.

The selectivity of musical advantage was also evident at higher perceptual levels. In the sequence recall task, listeners had to encode Cantonese tones and form representations in the familiarization phase. In the testing phase, listeners had to use Cantonese tones as memory cues to keep track of syllables. One may argue that the outperformance of musicians in the sequence recall task simply arose from enhanced encoding at the auditory level (as manifested in the AXB discrimination task), rather than higher level perceptual processes such as the formation of long-lasting representations and the use of Cantonese tones as memory cues. If enhanced sequence recall was simply due to enhanced encoding, one would expect that musicians outperform nonmusicians in recalling HL-ML sequences since musical advantage was evident in both HL and ML tone contexts in the AXB discrimination task. Critically, this prediction is contradictory to the present results. In this study, musicians did not outperform nonmusicians in recalling HL-ML sequences, even though musical advantage was found in the HL and ML tone contexts in the AXB discrimination task. More importantly, a clear musical advantage in LF-LR sequence recall was evident even though the musicians did not exhibit any relevant musical advantage in the AXB discrimination task. Taken together, results from the AXB and sequence recall tasks imply that the mechanism of music-to-language transfer is not that simple and straightforward. The OPERA hypothesis, in its current form, is a single-tier model that only accounts for how musical experience enhances speech perception at the auditory level. The current findings suggest the need to construct additional tier(s) in the OPERA hypothesis as the results have clearly shown that the musical advantage at higher perceptual levels cannot be fully accounted for by enhanced acoustic encoding at the auditory level.

In terms of methodological contribution, the selectivity of musical advantage identified herein highlights the need to include different tone contrasts or even pitch contrasts when studying cross-domain transfer. In some studies that tested musicians and nonmusicians on lexical tone discrimination, only overall accuracy (collapsed across different tones) was evaluated (e.g., Alexander et al., 2005). This might have masked the potential interactions between musical experience and tone context. Even within the music domain, a previous study showed that musical advantage was selective to pitch intervals among Cantonese listeners (Tong et al., 2018). Taken together, the present and previous findings encourage future studies to evaluate group differences in a fine-grained manner.

Building on the OPERA hypothesis (Patel, 2011), the present study suggests that the perceptual enhancement associated with musical experience is not general but specific to individual tones. This selectivity was evident in both low-level tone discrimination and high-level sequence recall. In contrast to the AXB task, musical experience facilitated the perception of a different set of Cantonese tones in the sequence recall task. This implied that the musicians' advantage in recalling tonal sequences could not simply be attributed to enhanced acoustic encoding. Future studies, preferably with larger sample sizes and enhanced statistical power, are warranted to elucidate the intricate mechanism governing music-to-language transfer.

Footnotes

1. According to the numerical notational system developed by Chao (1930), 5 represents the highest level of fundamental frequency (F0 hereafter) whereas 1 represents the lowest level of F0.

2. Studies on tone merging have shown that even native Cantonese listeners are not able to discriminate the HR-LR tone pair (e.g., Mok et al., 2013; Ou & Law, 2016). Similar to native Cantonese listeners, neither group in this study demonstrated above chance performance in discriminating the HR-LR tone pair. Thus, it was not included in the analysis.

Author Note

Correspondence concerning this article should be addressed to Dr. William Choi, Department of Early Childhood Education, The Education University of Hong Kong, 10 Lo Ping Road, Taipo, Hong Kong. E-mail: willchoi@eduhk.hk

References

Alexander, J., Wong, P. C. M., & Bradlow, A. R. (2005). Lexical tone perception in

musicians and non-musicians. Paper presented in *INTERSPEECH 2005 - Eurospeech,

9th European Conference on Speech Communication and Technology*, Lisbon,

Portugal.

Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W.

Strange (Ed.), *Speech perception and linguistic experience: Theoretical and

methodological issues in cross-language speech research* (pp. 167–200).

Timonium, MD: York.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception:

Commonalities and complementarities. In M. J. Murry & O. S. Bohn (Eds.),

*Language experience in second language speech learning: In honor of James Emil

Flege* (pp. 13–34). Amsterdam, The Netherlands: John Benjamins.

Burnham, D., & Mattock, K. (2007). The perception of tones and phones. In M. J. Munro &

O. S. Bohn (Eds.), *Second language speech learning: The role of language experience

in speech perception and production: In honor of James Emil Flege* (pp. 259–280).

Amsterdam, The Netherlands: John Benjamins.

Chao, Y. (1930). A system of tone-letters. *Le Maître Phonétique*, *45*, 24–27.

Choi, W., Tong, X., Gu, F., Tong, X., & Wong, L. (2017). On the early neural

perceptual integrality of tones and vowels. J*ournal of Neurolinguistics*, *41*, 11–23.

Choi, W., Tong, X., & Samuel, A. G. (2019). Better than native: Tone language

experience enhances English lexical stress discrimination in Cantonese-English

bilingual listeners. *Cognition*, *189*, 188–192.

Choi, W., Tong, X., & Singh, L. (2017). From lexical tone to lexical stress: A cross-language

mediation model for Cantonese children learning English as a second language.

Frontiers in Psychology, *8*, 492.

Cooper, A., & Wang, Y. (2012). The influence of linguistic and musical experience on

Cantonese word learning. *Journal of the Acoustical Society of America*, *131*, 4756-

4769.

Delogu, F., Lampis, G., & Olivetti-Belardinelli, M. (2006). Music-to-language

transfer effect: May melodic ability improve learning of tonal languages by native

nontonal speakers? *Cognitive Processing*, *7*, 203–207.

Dupoux, E., Peperkamp, S., & Sebastian-Galles, N. (2010). Limits on bilingualism

revisited: Stress 'deafness' in simultaneous French-Spanish bilinguals. *Cognition*,

114, 266–275.

Gandour, J. (1981). Perceptual dimensions of tone: Evidence from Cantonese. *Journal

of Chinese Linguistics*, *9*, 20–36.

Gu, W., & Lee, T. (2009). Effects of tone and emphatic focus on F0 contours of

Cantonese speech: A comparison with Standard Chinese. *Chinese Journal of

Phonetics*, *2*, 133–147.

Ho, Y.-C., Cheung, M.-C., & Chan, A. S. (2003). Music training improves verbal but

not visual memory: Cross-sectional and longitudinal explorations in children.

Neuropsychology, *17*, 439–450.

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of

    auditory skills. *Nature Reviews Neuroscience*, *11*, 599–605.

Lee. Y. S., Vakoch, D. A., & Wurm, L. H. (1996). Tone perception in Cantonese and

    Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, *25*,

    527–542.

Li, Y. (2016). English and Thai speakers' perception of Mandarin tones. *English Language*

    Teaching, *9*, 122–132.

Mok, P., Zuo, D., & Wong, P. (2013). Production and perception of a sound change in

    progress: Tone merging in Hong Kong Cantonese. *Language Variation and Change*,

    25, 341–370.

Nan, Y., Sun, Y., & Peretz, I. (2010). Congenital amusia in speakers of a tone language:

    Association with lexical tone agnosia. *Brain*, *133*, 2635–2642.

Ou, J., & Law, S. P. (2016). Individual differences in processing pitch contour and rise time

    in adults: A behavioral and electrophysiological study of Cantonese tone merging.

    Journal of the Acoustical Society of America, *139*, 3226–3237.

Paquette, S., & Goulet, G. M. (2014). Lifetime benefits of musical training. *Frontiers*

    in Neuroscience, *8*, 89.

Patel, A. D. (2011). Why would musical training benefit the neural encoding of

    speech? The OPERA hypothesis. *Frontiers in Psychology*, *2*, 142.

Qin, Z., & Mok, P. P. K. (2011, August 17). *Perception of Cantonese tones by Mandarin,*

English and French speakers. Paper presented at the 17th International Congress of Phonetic Sciences, Hong Kong, China.

Reid, A., Burnham, D., Kasisopa, B., Reilly, R., Attina, V., Rattanasone, N. X., & Best, C. T. (2015). Perceptual assimilation of lexical tone: The roles of language experience and visual information. *Attention, Perception, and Psychophysics*, *77*, 751–591.

Rodrigues, A. C., Loureiro, M. A., & Caramelli, P. (2010). Musical training, neuroplasticity and cognition. *Dementia and Neuropsychologia*, *4*, 277–286.

Schellenberg, E. G., & Mankarious, M. (2012). Music training and emotion comprehension in childhood. *Emotion*, *12*, 887–891.

Stoesz, B. M., Jakobson, L. S., Kilgour, A. R., & Lewycky, S. T. (2007). Local processing advantage in musicians: Evidence from disembedding and constructional tasks. *Music Perception*, *25*, 153–165.

Tillmann, B., Burnham, D., Nguyen, S., Grimault, N., Gosselin, N., & Peretz, I. (2011). Congenital amusia (or tone-deafness) interferes with pitch processing in tone languages. *Frontiers in Psychology*, *2*, 120.

Tong, X., Choi, W., & Man, Y. Y. (2018). Tone language experience modulates the effect of long-term musical training on musical pitch perception. *Journal of the Acoustical Society of America*, *144*, 690–697.

Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning*, *54*, 681–712.

Zheng, Y., & Samuel, A. G. (2018). The effects of ethnicity, musicianship, and tone

language experience on pitch perception. *Quarterly Journal of Experimental Psychology*, *71*, 2627–2642.

Table 1

Acoustic Parameters of all Male-produced Stimuli

Stimuli	Duration (ms)	F0 height (Hz)	Minimum F0 (Hz)	Maximum F0 (Hz)	F0 onset (Hz)	F0 offset (Hz)	F0 contour (Hz/ms)
ka-HL	596	155	151	168	168	158	-17
ka-HR	643	114	97	187	109	187	121
ka-ML	616	124	121	160	160	127	-54
ka-LF	580	92	89	104	104	89	-26
ka-LR	616	109	89	168	109	168	96
ka-LL	587	93	90	104	104	92	-20
ku-HL	591	168	165	175	175	167	-14
ku-HR	673	123	108	174	174	158	-24
ku-ML	700	133	129	148	138	148	14
ku-LF	659	106	94	127	127	108	-29
ku-LR	654	121	107	166	116	162	70
ku-LL	574	109	103	122	121	122	2
kɛ-HL	539	154	152	157	154	155	2
kɛ-HR	605	115	99	169	99	169	116
kɛ-ML	550	132	129	136	134	136	4
kɛ-LF	513	99	92	119	119	92	-53
kɛ-LR	600	111	94	168	114	168	90
kɛ-LL	531	102	97	117	117	98	-36
ki-HL	581	162	156	187	187	156	-53
ki-HR	644	122	110	176	119	176	89

ki-ML	758	134	127	160	160	145	-20
ki-LF	559	99	90	113	113	95	-32
ki-LR	700	121	106	164	115	164	70
ki-LL	679	113	111	130	130	119	-16
kɔ-HL	582	145	143	157	157	151	-10
kɔ-HR	632	115	103	152	138	150	19
kɔ-ML	566	125	122	127	127	126	-2
kɔ-LF	577	93	84	106	106	93	-23
kɔ-LR	661	118	99	177	102	177	113
kɔ-LL	655	105	100	127	127	113	-21

Note. HL = high level; HR = high rising; ML = mid-level; LF = low falling; LR = low rising;

LL = low level. All values are rounded off to the nearest integer.

Table 2

Acoustic Parameters of all Female-produced Stimuli

Stimuli	Duration (ms)	F0 height (Hz)	Minimum F0 (Hz)	Maximum F0 (Hz)	F0 onset (Hz)	F0 offset (Hz)	F0 contour (Hz/ms)
ka-HL	450	215	209	234	221	234	29
ka-HR	406	172	180	214	180	214	84
ka-ML	443	184	179	198	198	179	-43
ka-LF	553	154	76	191	190	77	-204
ka-LR	547	161	72	239	186	238	95
ka-LL	547	134	67	179	177	161	-29
ku-HL	452	230	123	249	249	225	-53
ku-HR	431	195	169	229	169	216	109
ku-ML	456	195	192	208	198	208	22
ku-LF	437	154	144	179	173	179	14
ku-LR	527	193	166	248	192	238	87
ku-LL	518	167	162	178	166	178	23
kɛ-HL	484	218	211	234	231	232	2
kɛ-HR	500	173	158	203	175	203	56
kɛ-ML	477	196	191	245	203	245	88
kɛ-LF	469	147	130	177	172	166	-13
kɛ-LR	454	166	152	206	169	206	82
kɛ-LL	602	162	153	186	169	185	27
ki-HL	463	231	228	240	229	240	24
ki-HR	483	151	84	222	84	218	277

ki-ML	503	179	171	207	178	206	56
ki-LF	403	162	152	184	166	184	45
ki-LR	548	188	165	233	179	232	97
ki-LL	569	168	163	186	170	186	28
kɔ-HL	463	224	219	239	239	237	-4
kɔ-HR	431	191	166	292	194	229	81
kɔ-ML	514	185	178	199	199	193	-12
kɔ-LF	479	148	76	199	198	199	2
kɔ-LR	537	172	148	215	184	211	50
kɔ-LL	584	77	75	86	77	86	15

Note. HL = high level; HR = high rising; ML = mid-level; LF = low falling; LR = low rising;

LL = low level. All values are rounded off to the nearest integer.

Table 3

Token-averaged Acoustic Parameters of the Six Cantonese Tones

Gender	Tone	Duration (ms)	F0 height (Hz)	Minimum F0 (Hz)	Maximum F0 (Hz)	F0 onset (Hz)	F0 offset (Hz)	F0 contour (Hz/ms)
Male	HL	578	157	153	169	168	157	-18
(N = 30)	HR	639	118	103	172	128	168	64
	ML	638	130	126	146	144	136	-11
	LF	578	98	90	114	114	95	-32
	LR	646	116	99	169	111	168	88
	LL	605	104	100	120	120	109	-18
Female	HL	462	224	198	239	234	234	0
(N = 30)	HR	450	176	151	232	160	216	121
	ML	479	188	182	211	195	206	22
	LF	468	153	116	186	180	161	-31
	LR	523	176	141	228	182	225	82
	LL	564	142	124	163	152	159	13
Averaged	HL	520	190	176	204	201	196	-9
(N = 60)	HR	545	147	127	202	144	192	93
	ML	558	159	154	179	170	171	5
	LF	523	125	103	150	147	128	-32
	LR	584	146	120	198	147	196	85
	LL	585	123	112	142	136	134	-3

Note. HL = high level; HR = high rising; ML = mid-level; LF = low falling; LR = low rising;

LL = low level. All values are rounded off to the nearest integer.

Table 4

Correlations between the Acoustic Parameters and Mean Accuracies of Musicians and

Nonmusicians across the Six Tone Contexts

Tone	Acc.	Duration	F0 height	Minimum F0	Maximum F0	F0 onset	F0 offset	F0 contour
HL	M	.40*	.42**	.28[‡]	.35*	.43**	.33*	.30[‡]
	NM	.16	.13	.11	-.04	.11	-.03	.27[‡]
HR	M	.07	.13	.08	.18	.24	.21	.28[‡]
	NM	-.18	.08	.04	-.01	.29[‡]	-.06	.01
ML	M	.08	-.15	-.13	-.20	-.07	-.29[‡]	.25
	NM	-.10	-.16	.10	-.16	-.15	-.16	.28[‡]
LF	M	.12	-.04	-.18	.06	-.03	.14	.40*
	NM	-.09	-.15	-.17	-.06	-.05	.01	.24
LR	M	.02	.11	.01	.15	.29[‡]	.19	.44**
	NM	.05	.18	.03	.18	.31[‡]	.17	.37*
LL	M	.01	.15	.01	.20	.24	.18	.40*
	NM	-.12	-.01	-.04	.01	.07	-.03	.12

Note. HL = high level; HR = high rising; ML = mid-level; LF = low falling; LR = low rising;

LL = low level; Acc. = mean accuracy; M = musician group; NM = nonmusician group; [**] p
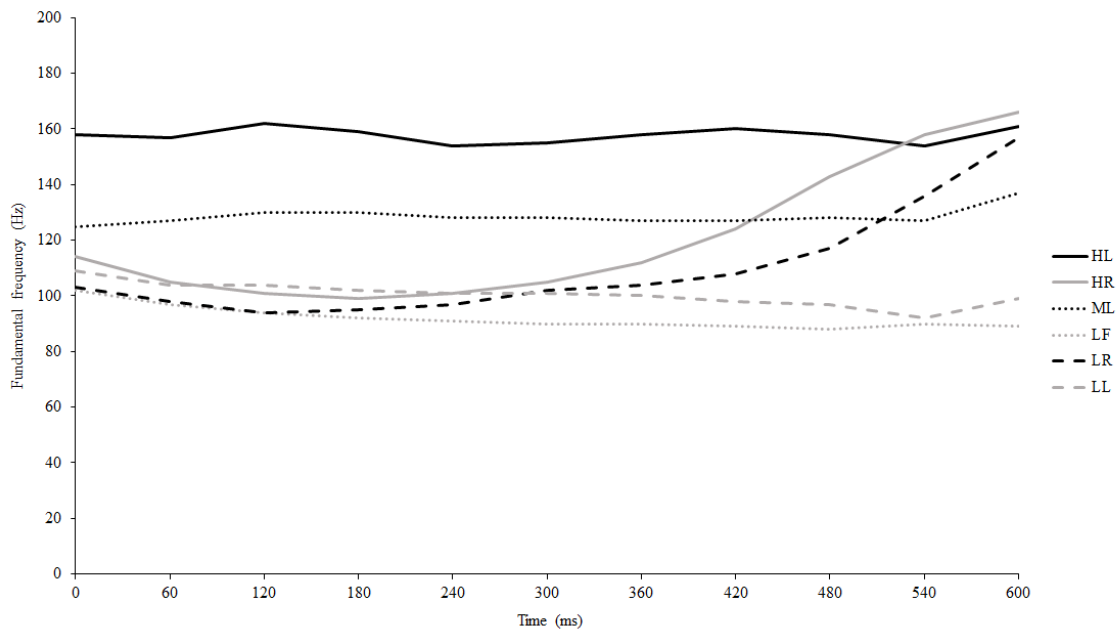
$< .01$; [*] $p < .05$; [‡] $p < .10$.

Figure 1. The fundamental frequency time graph of the six Cantonese tones, embedded in

/ta/, naturally produced by a male Cantonese speaker. HL = high level; HR = high rising; ML

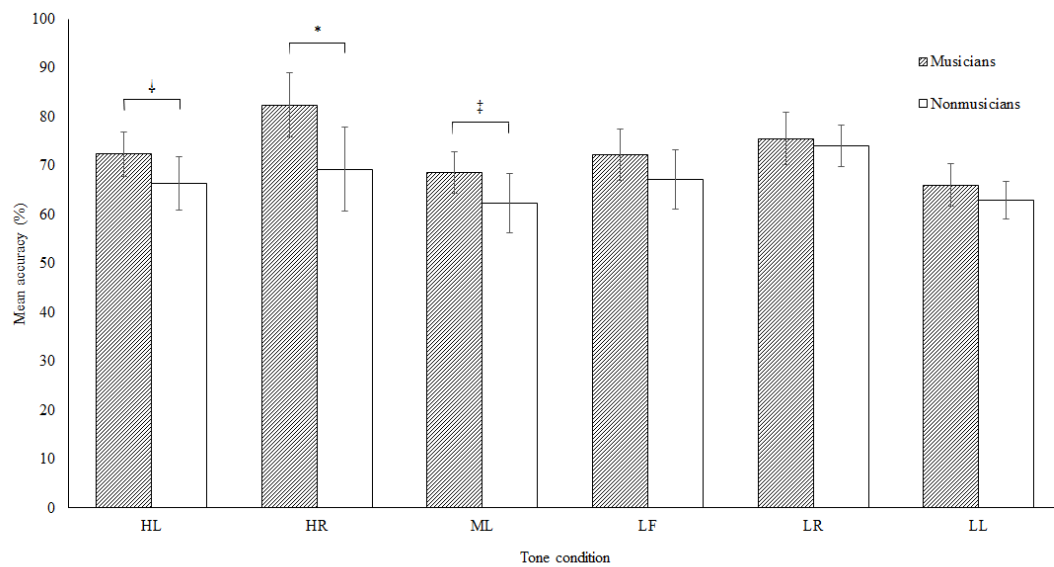= mid-level; LF = low falling; LR = low rising; LL = low level.

Figure 2. The mean accuracies of the musicians and the nonmusicians across the six Cantonese tone contexts. $^*p < .05$; $^\dagger p = .085$; $^\ddagger p = .082$. Error bars denote 95% confidence intervals.
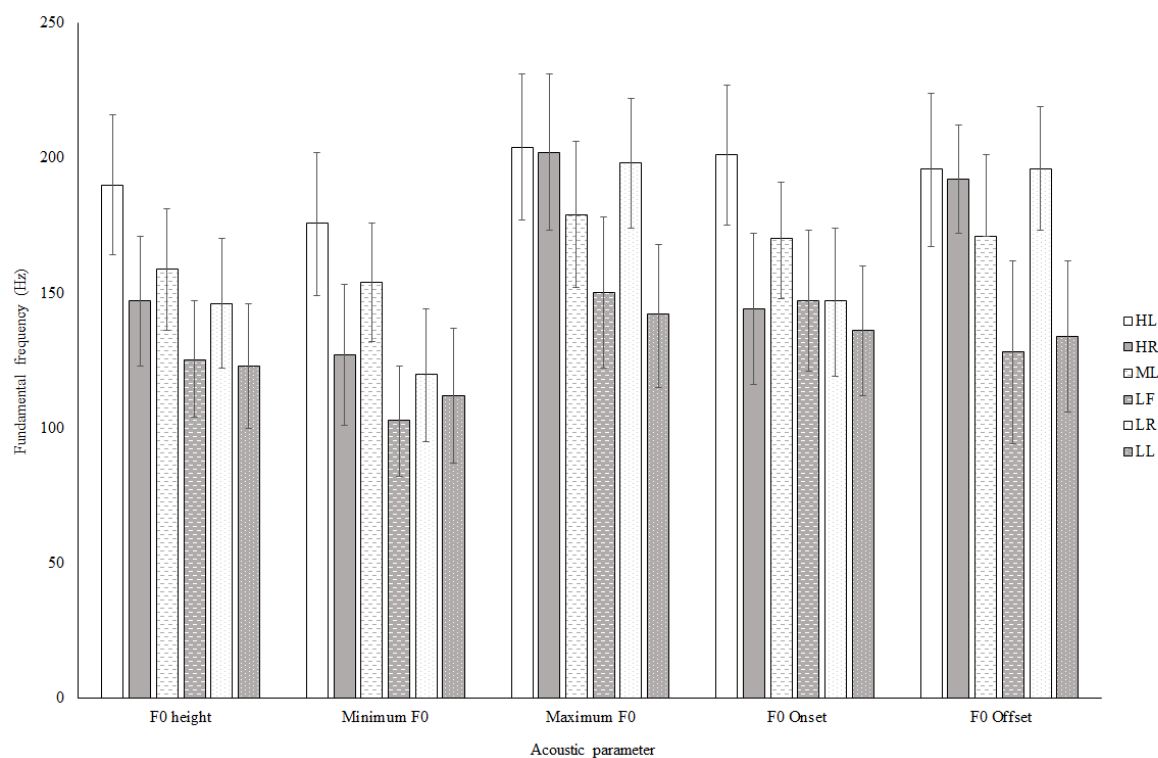
Figure 3. The average F0 heights, minimum F0, maximum F0, F0 onsets, and F0 offsets of

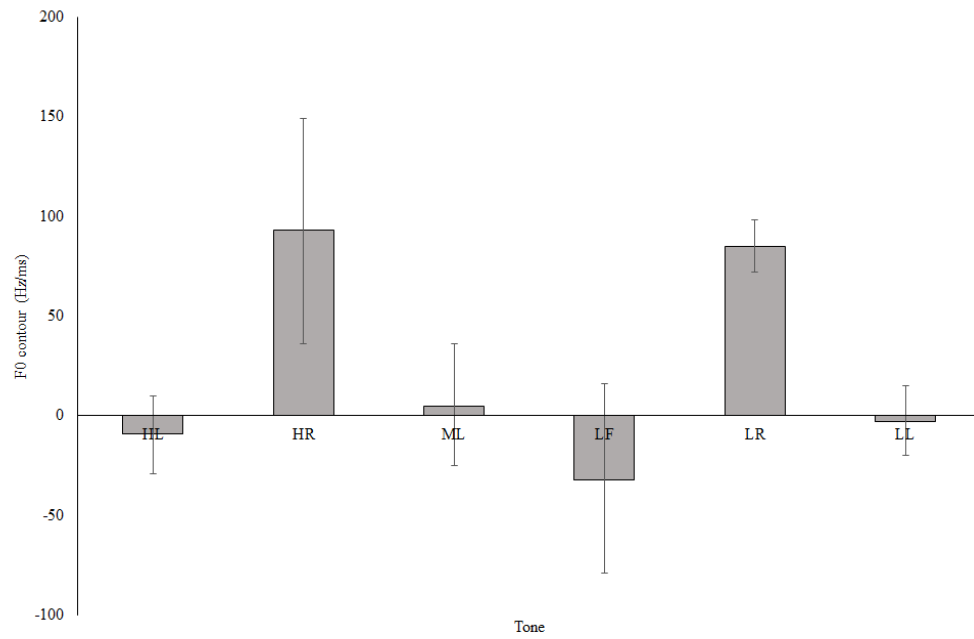the six Cantonese tones produced by both male and female speakers.

Figure 4. The average F0 contours of the six Cantonese tones produced by both male and
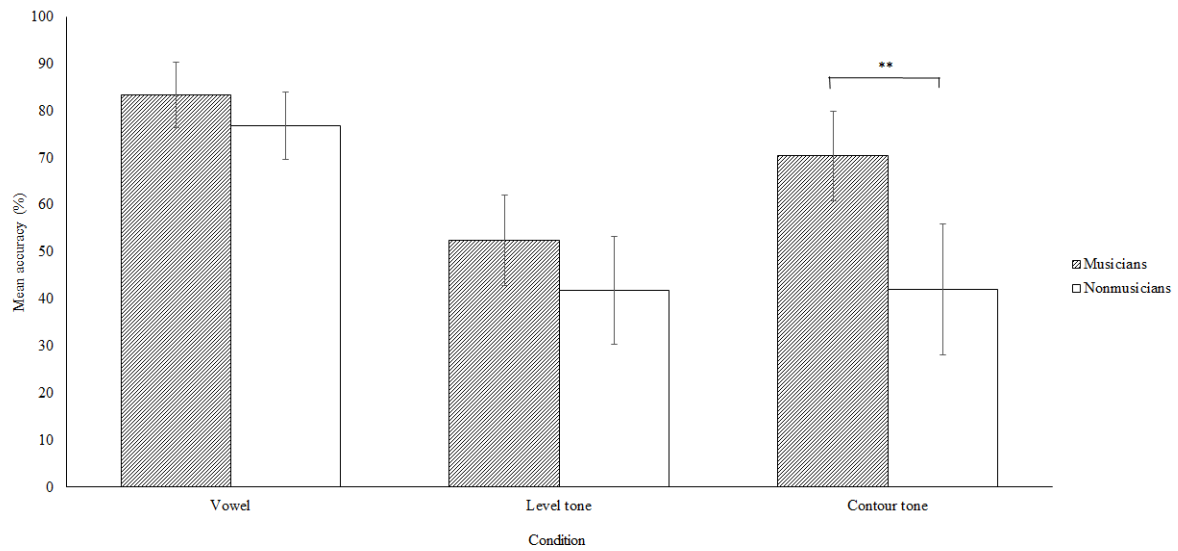
female speakers.

Figure 5. The mean accuracies of the musicians and the nonmusicians across the vowel, level

tone and contour tone contexts. **p < .01.