



Categorical Perception of Mandarin Pitch Directions by Cantonese-Speaking Musicians and Non-musicians

Si Chen^{1,2*}, Yike Yang¹ and Ratreë Wayland³

¹ Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong, SAR China, ² Hong Kong Polytechnic University-Peking University Research Centre on Chinese Linguistics, Hong Kong, SAR China, ³ Department of Linguistics, University of Florida, Gainesville, FL, United States

Purpose: This study is to investigate whether Cantonese-speaking musicians may show stronger CP than Cantonese-speaking non-musicians in perceiving pitch directions generated based on Mandarin tones. It also aims to examine whether musicians may be more effective in processing stimuli and more sensitive to subtle differences caused by vowel quality.

Methods: Cantonese-speaking musicians and non-musicians performed a categorical identification and a discrimination task on rising and falling continua of fundamental frequency generated based on Mandarin level, rising and falling tones on two vowels with nine duration values.

Results: Cantonese-speaking musicians exhibited a stronger categorical perception (CP) of pitch contours than non-musicians based on the identification and discrimination tasks. Compared to non-musicians, musicians were also more sensitive to the change of stimulus duration and to the intrinsic F_0 in pitch perception in pitch processing.

Conclusion: The CP was strengthened due to musical experience and musicians benefited more from increased stimulus duration and were more efficient in pitch processing. Musicians might be able to better use the extra time to form an auditory representation with more acoustic details. Even with more efficiency in pitch processing, musicians' ability to detect subtle pitch changes caused by intrinsic F_0 was not undermined, which is likely due to their superior ability to process temporal information. These results thus suggest musicians may have a great advantage in learning tones of a second language.

Keywords: categorical perception, musical experience, intrinsic F_0 , stimulus duration, tone

OPEN ACCESS

Edited by:

Chao-Yang Lee,
Ohio University, United States

Reviewed by:

Chang Liu,
University of Texas at Austin,
United States
Linjun Zhang,
Beijing Language and Culture
University, China

*Correspondence:

Si Chen
sarah.chen@polyu.edu.hk

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 24 May 2021

Accepted: 15 September 2021

Published: 14 October 2021

Citation:

Chen S, Yang Y and Wayland R (2021)
Categorical Perception of Mandarin
Pitch Directions by
Cantonese-Speaking Musicians and
Non-musicians.
Front. Psychol. 12:713949.
doi: 10.3389/fpsyg.2021.713949

INTRODUCTION

Pitch, the perceptual correlate of fundamental frequency (F_0), plays an important role in both music and language. Musicians and tone language speakers have in common enhanced sensitivity to small pitch changes associated with meaningful units, namely melodies for musicians and words for speakers of a tone language. While the effects of musicianship and experience with lexical tones

have been separately studied among native and non-native speakers of tone languages, respectively, the effects of musical experience on lexical tone perception among speakers of lexical tone languages have been rarely investigated. To fill this research gap, the current study investigates the effects of musical training and categorical perception (CP) of tones among tonal speakers. Potential factors such as stimulus duration and vowel quality are also explored.

This study aims to better understand the relationship between musical experience and linguistic processing by comparing native Cantonese musicians and non-musicians on their categorization of falling and rising pitch continua, representative of Mandarin tones. The continua are realized on a high [i] and a low vowel [a], associated with a high and a low intrinsic F_0 , respectively (Whalen and Levitt, 1995). Unlike Mandarin, Cantonese contrasts six lexical tone categories in open syllables (Xu and Mok, 2012). In addition, while the four Mandarin tones contrast with each other in terms of direction of F_0 movement (level, falling, and rising), three of the six tones in Cantonese tones differ in pitch height. Therefore, Cantonese listeners place more emphasis on average height of F_0 contour than do Mandarin listeners when processing tones (Peng et al., 2012). This may explain why, besides Mandarin tone 2 (high-rising) vs. tone 3 (low-falling-rising), Cantonese speakers have additional difficulties differentiating Mandarin tone 1 (high-level) from Mandarin tone 4 (high-falling) (Hao, 2012). However, due to their denser tonal system, native Cantonese listeners may be more efficient in processing pitch variation than native Mandarin listeners (Lee et al., 1996; Zheng et al., 2012).

Relationship Between Music and Language

Previous behavioral research has found that musical experience improves lexical tone perception among non-native tone speakers. For example, English-speaking musicians were significantly more skilled at identifying or discriminating Mandarin tones (Alexander et al., 2005; Lee and Hung, 2008) than English-speaking non-musicians, even when only partial F_0 information was present (Lee and Hung, 2008). Additionally, novice English musicians were more accurate and faster than non-musicians in their discrimination of Thai tones (Burnham et al., 2015).

Neurophysiological data has also revealed the facilitative effects of musicality on lexical tone processing. For instance, enhanced event-related brain potentials (ERPs) associated with Mandarin tone deviants were observed in French-speaking musicians compared to non-musicians (Marie et al., 2012). Furthermore, correlations between F_0 tracking quality and the amount of musical training and performance on identification and discrimination of Mandarin syllables were positive (Wong et al., 2007). Similarly, brainstem frequency following response (FFR) to homologs of musical intervals and of lexical tones showed that pitch tracking and pitch strength were more robust for English musicians compared to English non-musicians (Bidelman et al., 2011). The above findings indicated the overlap in music and language perception, suggestive of a common perceptual substrate for the two domains (Maggu et al., 2018). To account for how musical experience may benefit speech

processing, Patel (2011, 2012), and Patel (2014) proposed a neurocognitive model, OPERA, which stands for Overlap, Precision, Emotion, Repetition, and Attention. According to this model, speech processing benefits in musicians are attributed to overlapping brain networks engaged during speech and music perception.

In comparison to studies conducted on non-tone speaking musicians and non-musicians (Gottfried and Riester, 2000; Gottfried et al., 2004; Alexander et al., 2005; Gottfried, 2007; Wong and Perrachione, 2007; Lee and Hung, 2008) which largely revealed an overlap between music and language perception, the interaction between music and language among native tone speakers is not conclusive. For example, Tang et al. (2016) found increased neural activity in the discrimination of Mandarin tones (Tone 1 and Tone 2) and musical notes (C4 and G3 in piano timbre) among Mandarin-speaking musicians compared to Mandarin-speaking non-musicians. Similarly, Ong et al. (2020) found that Cantonese musicians were more accurate than non-musicians in their ability to discriminate and identify the most challenging tone pair (T23–T25). On the other hand, Thai musicians in Cooper and Wang (2012) were not superior to Thai non-musicians (or English-speaking musicians) in their ability to associate a change in pitch to a change in lexical meaning. Mok and Zuo (2012) found that, while French and English-speaking musicians outperformed non-musicians in their ability to identify Cantonese tones and their pure tone analogs, such musical advantage was not observed among native Cantonese listeners, suggesting that the linguistic and musical processing may belong to separate but overlapping domains, at least among native tone speakers.

In sum, behavioral and electrophysical studies have found that musical experience facilitates pitch perception among non-tone musicians. However, the facilitative effects of musical experience on pitch perception among native speakers of lexical tones remain inconclusive.

Musical Experience and Categorical Perception of Lexical Tones

Categorical perception, a classic paradigm in speech perception, refers to the ability to perceive and group linguistically distinct categories with equal physical changes along a continuum (Lieberman et al., 1957). The results of prior research suggested that tonal continua were largely perceived categorically by native tone listeners (Abramson, 1975; Wang, 1976; Francis et al., 2003), but continuously by non-native tone listeners (e.g., Hallé et al., 2004; Peng et al., 2010).

It has been reported that musical experience modulates CP of non-speech and speech F_0 continua (Zatorre and Halpern, 1979; Howard et al., 1992; Wu et al., 2015; Zhao and Kuhl, 2015a; Chen et al., 2020), suggesting that experience with the distinguishing of pitch categories defined along an F_0 continuum in the musical domain (i.e., musical notes) may induce CP of both linguistic and non-linguistic F_0 continua. For example, Zatorre and Halpern (1979) found that perception of a continuum of major and minor thirds whose component tones were sounded simultaneously was more categorical among musicians than among non-musicians.

Similarly, Howard et al. (1992) found that the ability to label members of a computer-synthesized continuum of major to minor was more categorical among the most musical compared to the least and the moderate musical listeners. Wu et al. (2015) compared identification and discrimination of Mandarin Tone 1–Tone 4 continuum by Mandarin musicians and non-musicians. While the steepness and the location of the category boundary as well as the between-category discrimination were comparable between the two groups, within-category discrimination was found to be enhanced among the musicians. The results suggested that musicality refines low-level auditory perception without interfering with higher-level, categorical processing of lexical tonal contrasts in native tonal listeners. Recently, Chen et al. (2020) found that perception of level-to-rising and level-to-falling pitch continua was more categorical among English-speaking musicians than among English-speaking non-musicians. Zhu et al. (2021) examined CP of Mandarin Tone 2–Tone 4 continuum and its non-speech, pure tone analogs, and found that mean amplitude of the mismatch negativities (MMNs) elicited by within-category deviants was significantly larger among amateur musicians than non-musicians for both types of continua. This result suggests that musical advantage extends to auditory processing of pitch at the pre-attentive level and is not confined to professional musicians. According to Bidelman (2017), musical training may improve the abilities of representing auditory objects and matching incoming sounds to memory templates, and these improved abilities may provide musicians with advantages in CP of speech.

However, musical training does not always promote or enhance CP among either non-native or native speakers of lexical tones. For example, Zhao and Kuhl (2015a) compared the perception of Mandarin Tone 2–Tone 3 continuum among English-speaking musicians, English-speaking non-musicians, and Mandarin-speaking non-musicians and found that in contrast to native Mandarin non-musicians, English-speaking musicians and non-musicians perceived the continuum non-categorically, and while short-term perceptual training improved perception, no evidence of categorical formation among either musicians or non-musicians post training. According to Bidelman et al. (2011), “Pitch encoding from one domain of expertise may transfer to another as long as the latter exhibits acoustic features overlapping those with which individuals have been exposed to from long-term experience or training” (p. 432). Zhao and Kuhl (2015b) compared musical pitch and lexical tone discrimination among Mandarin musicians, Mandarin non-musicians, and English musicians. No difference between Mandarin musicians and non-musicians was found in their sensitivity to lexical tones or in the pattern of within-pair sensitivity to the tone pairs. The English musicians showed significantly higher overall sensitivity to lexical tones than the two Mandarin groups and exhibited a different pattern of within-pair sensitivity, indicating that the processing of musical pitch and lexical pitch might be independent in nature. In addition, Chen et al. (2020) reported that Mandarin musicians did not consistently perceive rising and falling pitch directions more categorically than Mandarin non-musicians. A plausible

explanation is that perception of music and speech may implicate distinct processing mechanisms, and the processing of lexical tones makes use of other phonetic cues (e.g., duration and amplitude) besides F_0 (Liu and Samuel, 2004; Lee and Lee, 2010). Maggu et al. (2018)’s finding that Cantonese musicians did not differ from non-musicians on the brainstem encoding of lexical tones, but that they showed a more robust brainstem encoding of musical pitch as compared to non-musicians lends further support to the hypothesis that distinct mechanisms are engaged in the encoding of linguistic and musical pitch among native tone speakers.

To further probe the interaction between musical and linguistic training on pitch perception, we compared CP of Mandarin tones among Cantonese musicians and non-musicians.

The Role of Stimuli Duration and Vowel Quality in Perception of Tones

It has been found that perception of shorter vowels is more categorical than perception of longer vowels, suggesting the role of stimulus duration in CP. Duration is purported as one possible acoustic dimension affecting representation strength. According to the cue-duration hypothesis, acoustic information of consonants (e.g., formant transitions) is relatively short and less represented while information of formants in vowels is longer and better represented in auditory memory (Fujisaki and Kawashima, 1970). Moreover, the interaction between tones and the perceived vowel duration has been reported in several studies (Yu et al., 2014; Wang et al., 2017). For example, Yu et al. (2014) argues that perceived duration may be affected by F_0 slope and height. Usually, syllables with dynamic F_0 tend to be perceived as longer than those with flat F_0 . These results suggest an interaction between perceived duration and tonal shapes. Duration of stimulus also plays a critical role in pitch contour perception. It has been reported that for both native Mandarin and English speakers, the strength of CP increased as stimulus duration increased. In addition, native Chinese listeners showed stronger effects from stimulus duration in terms of category boundary sharpness, between- and within- category discrimination, and peakedness compared to native English listeners (Chen et al., 2017). These results are inconsistent with the cue-duration hypothesis stating that longer stimuli will be processed with weaker CP. They also revealed an influence of tone language background to the duration effect (Chen et al., 2017).

Also, a few studies reported enhanced processing of duration, both pre-attentively and attentively among musicians. For example, Marie et al. (2012) found larger pre-attentive and attentive responses to duration deviants among native speakers of Finnish, a language with phonemic vowel length contrast, and French musicians relative to non-musicians. In another study, Chobert et al. (2014) found that both the passive and the active processing of vowel duration and voice-onset-time (VOT) deviants were enhanced in musicians compared with non-musician children. These findings suggested that linguistic and musical expertise similarly influenced the processing of

pitch contour and duration in music and language, possibly because they tapped on the same pool of neural resources (Besson et al., 2011). The duration effect was mostly examined on native speakers or non-tonal speakers. Therefore, it is worth examining whether similar duration effect can be observed among musicians and non-musicians with a tone language background.

In addition to stimulus duration, intrinsic F_0 effects have been recently reported to contribute to CP (Chen et al., 2017). Intrinsic F_0 effects refer to the consistent correlation between F_0 -values and vowel height (Whalen and Levitt, 1995). In speech production, high vowels are correlated with higher F_0 -values and low vowels with low F_0 -values. But such correlation is reversed in speech perception, namely, high vowels are perceived to have a lower F_0 -value when they actually share the same F_0 (Wang et al., 1976; Stoll, 1984). Yu et al. (2014) also proposed a hypothesis that if perceptual compensation occurs for high vowels, where they are perceived to have a lower F_0 -value, then high vowels may in turn be perceived as longer. Chen et al. (2017) found that Mandarin and English listeners required a longer duration to perceive a tone on a low vowel than a high vowel. Chen et al. (2020) also reported that vowel quality significantly contributed to tone identification and sharpness of category boundary in English and Mandarin musicians. Vowel quality also plays a role in tone perception especially for musicians and they were better at teasing apart the factor of vowel quality that may affect the F_0 cue than non-musicians. The current study aims to examine the factor of vowel quality in tone perception by non-native tonal speakers with and without musical experience.

The Current Study

We have three research goals for the current study: (1) to investigate the effects of musical experience on CP of pitch in a non-native language by native speakers of a tone language; (2) to examine how longer stimulus duration affects auditory categorization of pitch among non-native tonal musicians and non-musicians; (3) to investigate if musicians are more sensitive to factors that may potentially influence pitch processing such as vowel quality and pitch directions than non-musicians.

METHODOLOGY

Participants

A total of 28 native speakers of Cantonese participated in the experiment. All participants speak Cantonese as the first and dominant language and they also speak English and Mandarin. Of all the participants, 14 were musicians (seven males, seven females; mean age \pm SD: 23.79 ± 2.40 ; age range: 19–27) and the other 14 were non-musicians (seven males, seven females; mean age \pm SD: 23.86 ± 2.70 ; age range: 19–27). The musicians began receiving formal musical training of western music instruments at an average age of 6.96 (± 1.95) years and all had regular practice of the instruments at the time of the experiment (mean years of musical experience \pm SD: 16.80 ± 3.31 ; range: 10–23). The non-musicians did not receive any after-school musical training. To confirm that the two groups of participants had similar background of Mandarin learning, we further collected information on the onset age of Mandarin learning, years of

Mandarin learning as well as self-reported proficiency of reading, writing, listening, and speaking abilities in Mandarin on a five-point scale as listed in **Table 1** (Point 1 indicates the lowest level of proficiency and point 5 indicates the highest level of proficiency). The participants reported no history of speaking, hearing, or language difficulty.

Stimuli

To examine the role of vowel quality, two sets of stimuli were created in the same way on low and high vowels [a] and [i]. A male native speaker of Mandarin with no reported speaking or hearing problem produced the Mandarin syllables [a] and [i] with the high-level Tone 1 using an Audio-Technica AT2020 microphone in a soundproof booth of the phonetics lab at the University of Florida. For each set of stimuli, the pitch contour of the original target syllable was manipulated with the pitch synchronous overlap add (PSOLA) method (Moulines and Laroche, 1995) in Praat (Boersma and Weenink, 2015). Our stimuli were all linear, and the slope and intercept parameters followed the estimates of a previous study based on a corpus of Mandarin speech (Prom-on et al., 2009). According to Prom-on et al. (2009), the slope and intercept for the rising Tone 2 in Mandarin are 93.4 and -2.2 st, respectively. Equation (1) below was used to transform st from Hertz (Hz):

$$\text{Number of st} = \frac{12}{\log_{10} 2} * \log(F_{02}/F_{01}) \quad (1)$$

where F_{01} and F_{02} represent the lower F_0 and the higher F_0 , and the number of st measures the distance between F_{01} and F_{02} in Hz (Lehnert-LeHouillier, 2013). Following Xu et al. (2006), we set the F_{02} -value at 130 Hz, and calculated the F_{01} -value based on the chosen F_{02} -value and the intercept value of -2.2 st.

For each vowel, nine durations were manipulated: 200, 180, 160, 140, 120, 100, 80, 60, and 40 ms. In total, there were 18 continua (9 duration * 2 vowels). The stepwise onset values with different duration values for the rising tone are as shown in **Table 2** and those for the falling tone are listed in **Table 3**. **Figures 1, 2** are examples of two rising continua with the durations of 200 and 40 ms.

Using a rising continuum as an example, we describe how steps in a continuum are calculated as follows. First, the offset

TABLE 1 | Participants' background information.

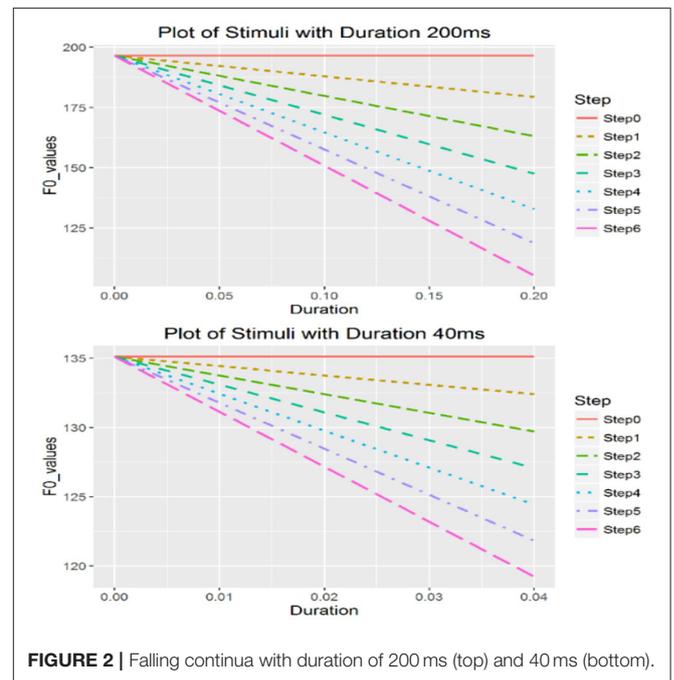
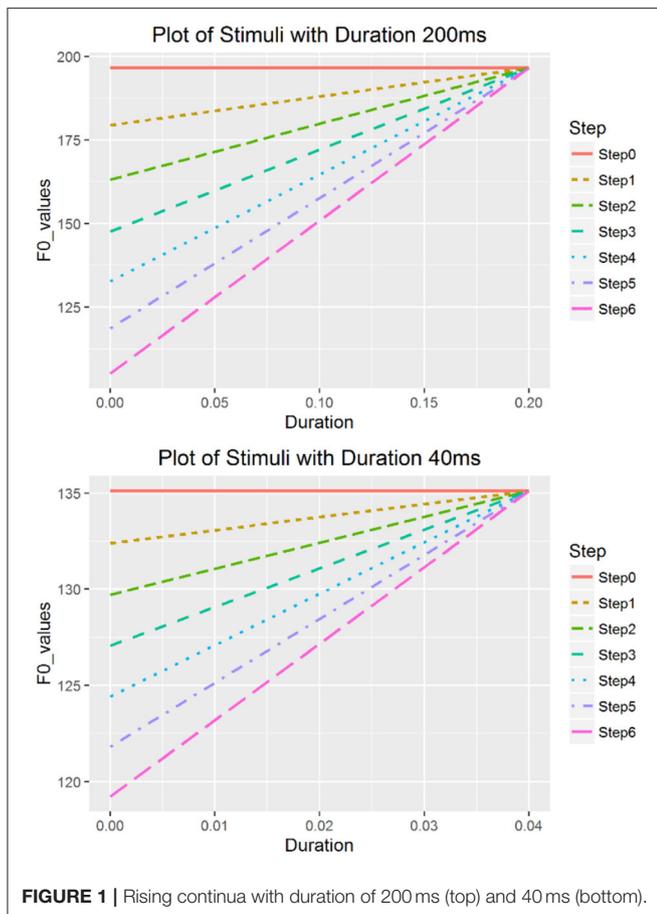
Factor	Musician	Non-musician
Age	23.79 \pm 2.40	23.86 \pm 2.70
Onset age of musical training	6.96 \pm 1.95	N/A
Years of musical training	16.80 \pm 3.31	N/A
Onset age of Mandarin learning	5.43 \pm 2.26	5.79 \pm 1.82
Years of Mandarin learning	7.85 \pm 2.56	8.14 \pm 2.20
Self-reported reading ability in Mandarin	4.57 \pm 0.49	4.71 \pm 0.45
Self-reported writing ability in Mandarin	4.43 \pm 0.73	4.36 \pm 0.48
Self-reported listening ability in Mandarin	3.64 \pm 1.11	3.71 \pm 0.45
Self-reported speaking ability in Mandarin	3.64 \pm 0.72	3.50 \pm 0.50

TABLE 2 | Onset values for each step varied by duration for linear rising pitch directions.

Duration	0.2	0.18	0.16	0.14	0.12	0.1	0.08	0.06	0.04
Step 0	196.56	187.56	178.97	170.78	162.96	155.50	148.38	141.59	135.11
Step 1	179.43	172.62	166.09	159.84	153.86	148.14	142.66	137.42	132.40
Step 2	163.12	158.31	153.69	149.26	145.02	140.94	137.04	133.30	129.72
Step 3	147.58	144.61	141.76	139.03	136.41	133.91	131.52	129.24	127.06
Step 4	132.77	131.49	130.27	129.13	128.05	127.04	126.11	125.24	124.43
Step 5	118.67	118.92	119.21	119.55	119.92	120.33	120.79	121.28	121.82
Step 6	105.23	106.89	108.57	110.28	112.01	113.78	115.57	117.39	119.24

TABLE 3 | Offset values for each step varied by duration for linear falling pitch directions.

Duration	0.2	0.18	0.16	0.14	0.12	0.1	0.08	0.06	0.04
Step 0	196.56	187.56	178.97	170.78	162.96	155.50	148.38	141.59	135.11
Step 1	179.43	172.62	166.09	159.84	153.86	148.14	142.66	137.42	132.40
Step 2	163.12	158.31	153.69	149.26	145.02	140.94	137.04	133.30	129.72
Step 3	147.58	144.61	141.76	139.03	136.41	133.91	131.52	129.24	127.06
Step 4	132.77	131.49	130.27	129.13	128.05	127.04	126.11	125.24	124.43
Step 5	118.67	118.92	119.21	119.55	119.92	120.33	120.79	121.28	121.82
Step 6	105.23	106.89	108.57	110.28	112.01	113.78	115.57	117.39	119.24



values for each duration were calculated with Equation (2):

$$X(t) = 93.4 * t - 2.2 \tag{2}$$

where t is the duration in s, and $X(t)$ stands for the semitone (st)-values at the offset of the tone. For example, for the 200 ms

continuum, $t = 200$ ms (0.2 s) and the onset can be calculated by setting $t = 0$ s, which is -2.2 st (or 123.02 Hz); and the offset can be calculated by setting $t = 0.2$ s, which is 16.48 st (or 196.56 Hz). Therefore, the onset-to-offset distance [i.e., $\Delta X(t)$] is $16.48 - (-2.2) = 18.68$ st in this case. Stimuli with various onsets were created based on the calculated intercept value of -2.2 st, which was also the cutoff point. The extreme points of onset values were then determined so that the distance between the lowest onset value (-8.43 st or 105.23 Hz) and -2.2 st was one third of the distance between the highest onset value (16.48 st or 196.56 Hz) and -2.2 st. The highest onset was defined as the same value as the obtained offset value (196.56 Hz) and is used to generate a level tone with equal onset and offset values. After obtaining the highest and lowest onsets, steps were created between them based on the ERB scale instead of Hertz because the former reflects natural perception (Xu et al., 2006). Seven stimuli with equal perceptual distance were created for each duration value in a continuum, and the onset values were then transformed back into Hertz.

Our resynthesizing procedure is similar to Peng et al. (2010): (1) we adjusted the duration of the stimuli to the duration values in **Tables 1, 2**; (2) we peak normalized the stimuli to the same intensity level; (3) We adjusted the pitch points according to the values in **Tables 1, 2**. The set of stimuli used by Chen et al. (2020) was used in this current study.

For the identification task, stimuli with a rising pitch continuum were grouped in one block and those with a falling pitch continuum were grouped in another block. There were 630 stimuli (5 repetitions * 7 steps * 9 duration * 2 syllable) in each block.

Since a one-step difference is too difficult to perceive (Francis et al., 2003), so this study used two-step difference pairs for the same-difference discrimination task. Again, the stimulus presentation was blocked by rising and falling pitch directions. Within each block, the different pairs were presented in either the forward order (0-2, 1-3, 2-4, 3-5, 4-6) or the backward order (2-0, 3-1, 4-2, 5-3, 6-4). All the same and different pairs were repeated twice in the task. Overall, there were 612 trials in each block [9 durations * (7 same pairs + 10 different pairs) * 2 syllables * 2 repetitions].

Procedure

All the participants signed informed consent forms in compliance with a protocol approved by the Human Subjects Ethics Sub-committee at the Hong Kong Polytechnic University and participated in the experiment with a GMH C 8.100 D headset at the Speech and Language Sciences Lab of the Hong Kong Polytechnic University. An identification task and a same-difference discrimination task were implemented in E-Prime 2.0 (Schneider et al., 2012). There was one practice block and two test blocks for each task where the practice block always preceded the test block. Within each block, the order of stimulus presentation was randomized. The order of block presentation for each participant was also counterbalanced and the block order was kept identical for the musician group and the non-musician group.

Training and Practice

Prior to the actual tasks, the participants were first familiarized with the tasks and the participants were required to identify the pitch directions they heard by pressing the number key 1 for a level tone and the number key 2 for a rising or falling tone in a practice session of the identification task. In the practice session for the same-difference discrimination task, the participants listened to a pair of stimuli at a time and were asked to decide whether the stimuli had the same or different pitch direction by pressing the number key 1 for “the same” and the number key 2 for “different.” Only stimuli with the longest duration value (0.2 s) at the two endpoints of each continuum were used for the practice sessions. A minimum threshold [75% (12 out of 16 trials); 71% (17 out of 24 trials) for the identification and the discrimination tasks, respectively] of correct responses was set for each practice session to make sure that the participants were able to finish the tasks and can identify or discriminate the pitch directions with either a level or steep rising/falling tone with the longest duration. Only those who passed the threshold proceeded to the experimental tasks. The reasons for including the practice session is that we need to make sure that the participants are familiar with the procedure required by the tasks and are able to press the right keys when listening to the pitch directions.

The Identification Task

In the identification task, the participants followed the same procedure as in the practice session. Once they heard a stimulus, they needed to decide whether it was a level tone or a rising/falling tone by pressing “1” for the former and “2” for the latter at a self-paced rate. The next stimulus was presented automatically after a response was given.

The Same-Different Discrimination Task

In the discrimination task, the participants listened to a pair of stimuli in each trial and were asked to decide whether the two stimuli were the same or different in terms of pitch direction at a self-paced rate. There were 34 trials of stimuli for each duration value [(7 same pairs + 10 different pairs) * 2 repetitions]. The 34 trials were divided into five two-step comparison units: 0-2, 1-3, 2-4, 3-5, and 4-6, where each unit had four types of comparisons: AA, AB, BA, and BB. The same trials were included. For example, the 3-3 pair was included in both 1-3 and 3-5 units. Therefore, there were eight trials (4 types of comparison * 2 repetitions) for each unit. In addition, the ISI of 500 ms was used as previous research suggests that 500 ms is the time needed to maximize the differences in between- vs. within-category discrimination (Pisoni, 1973; Xu et al., 2006; Peng et al., 2010).

The stimuli were presented automatically after a response was given. For further analyses, d' -prime (d') scores were computed from raw discrimination responses with the Equation (3):

$$d' = z(H) - z(F) \quad (3)$$

where d' is the d' -prime score, H is the hit rate (i.e., “different” response given for “different” trials), F is the false alarm rate (i.e., “different” response given for “same” trials) and z is z -transform (Creelman and Macmillan, 2004).

Data Analysis

There are three features of CP, namely a sharp category boundary, a discrimination peak, and prediction of discrimination from identification. The data obtained were, therefore, analyzed to examine the effects of musical background (musicians vs. non-musicians), pitch direction (rising vs. falling), vowel quality (low vs. high), and duration (nine different duration values) on these characteristics of CP.

The Identification Task

The identification data was analyzed to see if category boundary sharpness and category boundary location were affected by pitch direction type (falling vs. rising), vowel quality (low [a] vs. high [i]), stimulus duration (nine values: 40–200 in 20 ms increment) and musicianship (musicians vs. non-musicians). To achieve these goals, a generalized linear mixed model with subjects as a random effect was fitted to the data using the lme4 package (Bates et al., 2015) in R (R Core Team, 2018).

To perform the analyses, the data were divided into eight subgroups based on musical background, pitch direction, and vowel quality: FCMA, FCMI, RCMA, RCMI, FCNA, FCNI, RCNA, and RNI, where F and R represent falling and rising pitch directions, CM and CN stand for Cantonese musicians and non-musicians, and A and I for [a] and [i] syllables. For example, RCNA is the subset of data which includes rising pitch direction (R) on the syllable [a] (A) identified by Cantonese non-musicians (CN). Within each subgroup, a generalized linear mixed model was fitted with identification scores (0 or 1) as the response variable (the accuracy rate can be calculated from 0 and 1 s) and step number ($x = 0-6$) as a factor. The model is similar to a logistic regression model when only the fixed effects are considered in Equation (4).

$$\log_e \left(\frac{p_1}{1 - p_1} \right) = b_0 + b_1 x \quad (4)$$

In this equation, the coefficient b_1 stands for the sharpness of category boundary. To perform a *post-hoc* analysis on the effects of sharpness of boundary by duration, we carried out pairwise comparisons between different duration values for each of the eight subgroups. Specifically, we fit a model treating the coefficient b_1 from each pair of duration values as the same and conducted a likelihood ratio test to compare it to another model that treats them as different. Significant differences between the two models would indicate significant differences of the coefficient b_1 between two duration values. We also tested whether musical background, pitch direction and vowel quality would influence the values of b_1 . In addition, we modeled the relationship between the sharpness of category boundary and duration for musicians and non-musicians.

For category boundary location, after obtaining the estimates for b_0 and b_1 , $P_1 = 0.5$ was used to estimate the step number at category boundary within musician and non-musician groups, as shown in Equation (5):

$$x_{cb} = -b_0/b_1 \quad (5)$$

Once the individual category boundary was obtained for subjects in each subgroup, a linear mixed effects model was fitted, with pitch direction, musicianship, duration, and vowel quality as fixed effects and subjects as random effects, followed by *post-hoc* analyses. Linear regression models were fitted separately for musicians and non-musicians to test the relationship between duration and category boundary.

Finally, formulas for minimum stimulus duration required to perceive a rising or a falling F_0 from a level F_0 were derived using Equation (6):

$$t = b_0 + b_1 d \quad (6)$$

where t is the duration needed to perceive d st differences from level tones. For each duration value, the estimated step (identification rate equaled 0.5) was recorded and set as a cut-off point. Step numbers smaller than this point indicated that the stimulus was more likely to be identified as a level tone. The step number was transformed back to st-values with respect to the baseline (level tone as step 0) for each duration value. Linear mixed effects models were then fitted to obtain a relationship between cut-off st-values and duration, and it was assumed that the cut-off st-values were the smallest values for a rising or falling pitch direction to be perceived as different from a level tone. Minimum stimulus duration formulas for each of the eight subgroups were derived separately, and more general formulas for musicians and non-musicians to perceive each pitch direction were also obtained.

The Discrimination Task

d -prime's scores based on correct (hits) and incorrect (false alarms) responses were obtained in the discrimination task, and a generalized linear mixed model were fitted with duration, musical background, and vowel quality and all the two-way and three-way interactions as fixed factors and subject as a random factor.

To explore the relationship between between- and within-category discrimination, d -prime scores of between- and within-category discrimination for each subgroup according to its category boundary was calculated (Wu et al., 2015). The d -prime scores were calculated for all nine duration values. When the category boundary was <1 or >5 , the d -prime score was not calculated, because the steps were constrained between 0 and 6 steps. Linear mixed-effects models were then fitted to examine the contribution of musical background, pitch direction, vowel quality, and duration to the d -prime scores, followed by *post-hoc* analyses. Pairwise comparison of duration was performed, and linear regression models were fitted to examine the relationship between discrimination scores and duration.

The peakedness of discrimination function was estimated from the difference between P_{bc} and P_{wc} . P_{bc} (between-category discrimination) is defined as P of the comparison unit corresponding to the category boundary, and P_{wc} (within-category discrimination) is defined as the average of two comparison units at the extremes of the continuum (P_{02} and P_{46}) (Pisoni, 1973). Linear models were fitted to examine whether there were any significant contributing factors. Pairwise

comparison of duration with respect to peakedness was conducted, and regression models were also fitted to investigate the relationship between peakedness and duration.

Predicting Discrimination From Identification

According to Pollack and Pisoni (1971), the predicted discrimination score P^* can be calculated by Equation (7):

$$P^* = [1 + (P_A - P_B)^2] / 2 \quad (7)$$

where P_A and P_B are the identification scores in a comparison unit. This equation assumes that the discrimination can be solely determined by the identification of the two stimuli A and B . The correlation between the predicted and the observed discrimination scores for each comparison subgroup was calculated based on different trials of stimuli A and B by optimizing linear regression models after Fisher's z transformation. Next, the effects of musical background, pitch direction, vowel quality, and duration on the correlation was tested. The mean difference between the predicted and observed discrimination scores $P - P^*$ were also calculated. The optimized linear model was selected based on the stepwise optimization algorithm using the function "step" (Hastie and Pregibon, 1992) to model the relationship between the distance and the tested variables.

RESULTS

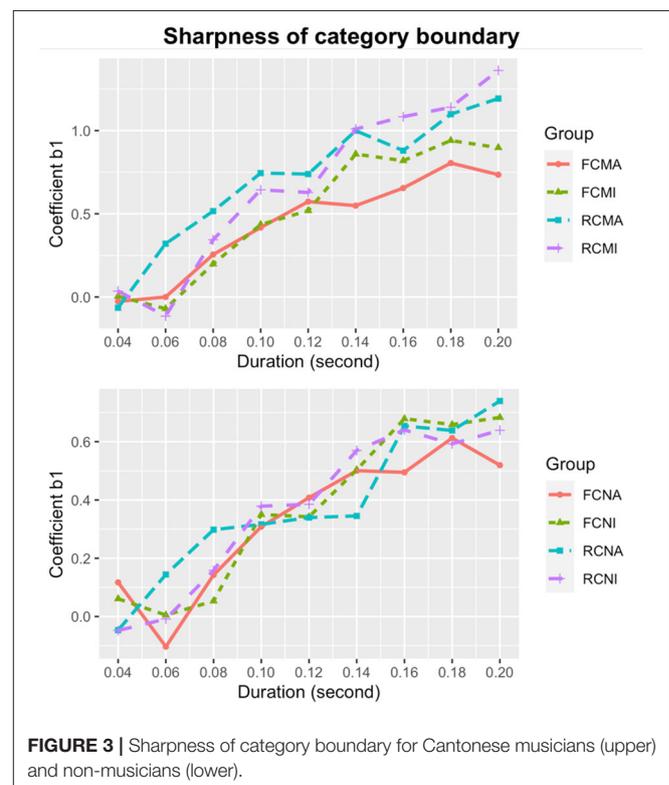
Identification

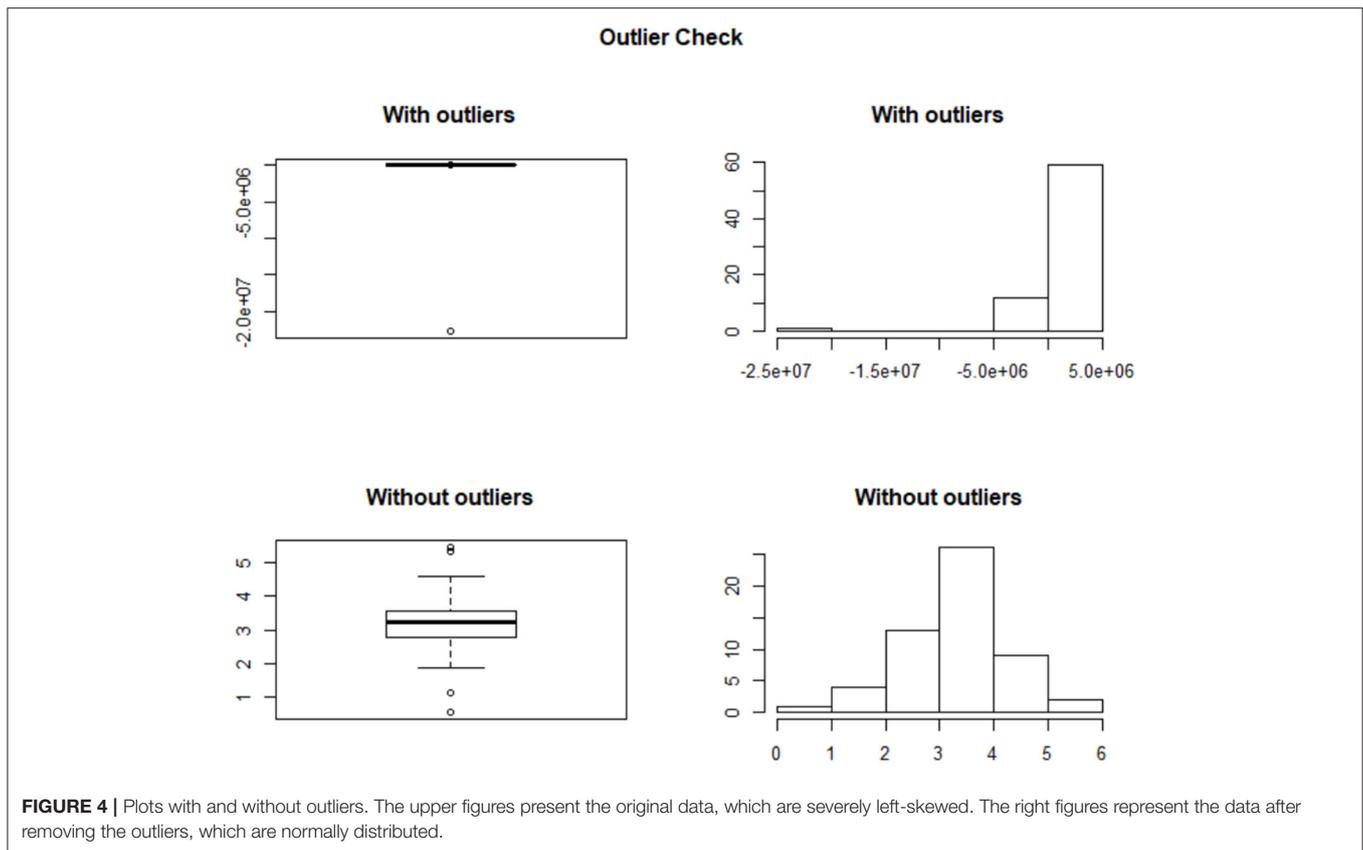
A generalized linear mixed effects model was fitted to the pitch direction (i.e., level vs. rising/falling) identification data obtained from Cantonese listeners with and without musical experience. The dependent variable was response rate, and the independent variables included vowel quality, pitch direction, duration, musical experience, and their interactions. The results revealed significant main effects of vowel quality [$\chi^2_{(1)} = 26.503$, $p < 0.001$] and pitch direction [$\chi^2_{(1)} = 104.9$, $p < 0.001$] as well as a marginally significant main effect of duration [$\chi^2_{(1)} = 3.285$, $p = 0.07$]. The main effect of musical experience, however, did not reach significance [$\chi^2_{(1)} = 1.017$, $p = 0.313$]. These results indicated that a syllable with the high vowel [i] was more likely to be identified as a level tone than a syllable with the low vowel [a], and a rising syllable was more likely to be identified as a level than a falling syllable. The chance of a syllable being identified as a level tone also decreased with the increase of syllable duration. The two-way interactions were significant for vowel quality and musical experience [$\chi^2_{(1)} = 9.356$, $p = 0.002$], musical experience and pitch direction [$\chi^2_{(1)} = 21.465$, $p < 0.001$], and musical experience and duration [$\chi^2_{(8)} = 169.99$, $p < 0.001$]. *Post-hoc* tests suggested that (1) vowel quality did not influence the identification of non-musicians, but musicians were more likely to identify a high vowel /i/ as a level tone; (2) both groups of participants were more likely to identify a falling syllable as a level tone; (3) with the increase of duration, musicians tended to identify a syllable as a rising/falling tone, while non-musicians tended to identify a syllable as a level tone.

Sharpness of Category Boundary

The estimates of the coefficient b_1 (sharpness of category boundary) of all nine stimulus durations for each subgroup are plotted in **Figure 3**. From **Figure 3**, it is evident that category boundary becomes sharper as the stimulus duration increases for all subgroups. In addition, overall, the musicians showed a sharper category boundary than the non-musicians.

We extracted all the values of category boundary sharpness b_1 and treat it as the dependent variable. The independent variables included vowel quality, pitch direction, duration, musical experience, and their interactions. Likelihood ratio tests revealed significant effects of musical experience [$\chi^2_{(1)} = 97.687$, $p < 0.001$]; pitch direction [$\chi^2_{(1)} = 26.733$, $p < 0.001$]; and vowel quality [$\chi^2_{(1)} = 26.503$, $p < 0.001$]. These results suggested that, on average, category boundary was significantly sharper for musicians than non-musicians [$b_1 = 0.587$ vs. 0.363]; for the rising pitch than the falling pitch direction [$b_1 = 0.536$ vs. 0.414], and for the high vowel /i/ than the low vowel /a/ [$b_1 = 0.483$ vs. 0.483]. There were also significant interactions between musical experience and pitch direction [$\chi^2_{(3)} = 150.47$, $p < 0.001$], between musical experience and vowel quality [$\chi^2_{(3)} = 110.48$, $p < 0.001$], and between musical experience and duration [$\chi^2_{(10)} = 3686.4$, $p < 0.001$]. For both groups of listeners, follow up tests revealed higher b_1 (sharper category boundary) values for the rising than the falling pitch directions and for the high vowel [i] than the low vowel [a]. However, none of the differences reached statistical significance.





To examine the significant interaction between musicianship and stimulus duration on category boundary sharpness, regression models were separately fitted for musicians and non-musicians. For the musicians, a regression model with an extra quadratic term did not significantly improve the model compared to a linear regression model, as suggested by a likelihood ratio test [$\chi^2_{(2)} = 0.187, p = 0.911$]. The linear regression model in Equation (8) was thus selected and it captured the relationship between the duration and sharpness well [$F_{(1,34)} = 153.1, p < 0.001$; adjusted $R^2 = 0.813$]. For the non-musicians, a regression model with an extra quadratic term did not significantly differ from a linear regression model [$\chi^2_{(2)} = 2.214, p = 0.331$], so the linear model in Equation (9) was adopted [$F_{(1,34)} = 251.7, p < 0.001$; adjusted $R^2 = 0.878$].

$$b_1 = 7.067 * d - 0.261 \quad (8)$$

$$b_1 = 4.498 * d - 0.177 \quad (9)$$

The formulae show that the sharpness of category boundary for both Cantonese musicians and non-musicians increases with the stimulus duration, but the musicians exhibit a steeper slope and thus have a faster increment of sharpness of category boundary with increased duration.

Category Boundary Location

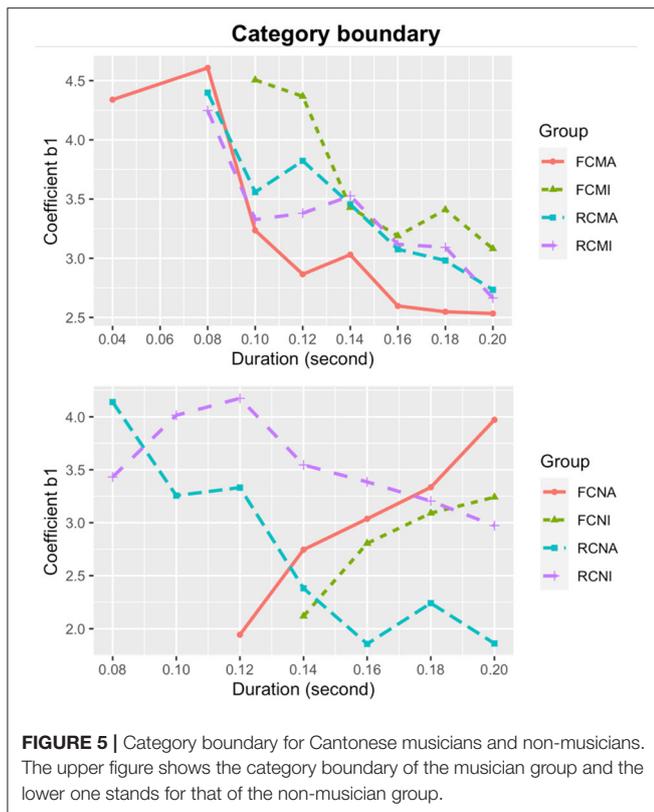
Category boundary for each subgroup and each duration value was calculated based on the estimated coefficients b_0 and b_1 from

the generalized linear models. After removing all the outliers¹ as illustrated in **Figure 4**, we plotted the category boundary against stimulus duration for each subgroup in **Figure 5**.

In general, we see that category boundary shifted to smaller values as the stimulus duration increased. A linear regression model was fitted, where the dependent variable was category boundary, and the independent variables included musicianship, vowel quality, duration, and pitch direction. A significant effect of stimulus duration ($p < 0.001$) was found, but a marginal effect of musical experience was identified ($p = 0.068$), wherein boundary locations were smaller for non-musicians than musicians. No effects of vowel quality ($p = 0.143$) or pitch direction ($p = 0.846$) were found.

To capture the relationship between category boundary location and stimulus duration for Cantonese musicians and non-musicians, regression models were fitted to the data. For the musicians, no significant difference was observed between a regression model with only the slope and intercept terms and a regression model with an extra quadratic term, as suggested by a likelihood ratio test [$\chi^2_{(1)} = 1.793, p = 0.181$], so the simple model [$F_{(1,24)} = 33.77, p < 0.001$; adjusted $R^2 = 0.567$] was adopted [Equation (10)]. Similarly, for the Cantonese non-musicians, no significant difference was found between a simple

¹Following the Tukey's method, outliers are defined as values above and below the $1.5 * IQR$ (interquartile). The script for detecting and removing the outliers can be found at: <https://goo.gl/4mthoF>.



model and a complex model [$\chi^2_{(1)} = 2.328, p = 0.127$], so the simple model [$F_{(1,19)} = 0.495, P = 0.490$; adjusted $R^2 = -0.026$] was adopted [Equation (11)].

$$cb = -12.008 * d + 5.068 \quad (10)$$

$$cb = -3.409 * d + 3.509 \quad (11)$$

We can infer from the equations that the category boundary shifted to smaller values as the stimulus duration increases, and the decrease rate is faster for the musician group.

Same-Different Discrimination Task

The raw data of the same-different discrimination task were transformed into a response of either 0 (same) or 1 (different) to fit generalized linear mixed effects models to test the main effects and interaction terms. The dependent variable was the response variable, and the independent variables included musicianship, vowel quality, duration, and pitch direction and their interactions. The results revealed significant main effects of vowel quality [$\chi^2_{(1)} = 6.451, p = 0.011$], pitch direction [$\chi^2_{(1)} = 583.31, p < 0.001$], and duration [$\chi^2_{(1)} = 2507.2, p < 0.001$]. The main effect of musical experience did not reach significance [$\chi^2_{(1)} = 2.377, p = 0.123$]. The two-way interactions were significant for vowel quality and musical experience [$\chi^2_{(2)} = 27.002, p < 0.001$], vowel quality and pitch direction [$\chi^2_{(2)} = 10.417, p = 0.005$], vowel quality and duration [$\chi^2_{(1)} = 4.867, p = 0.027$], musical experience and pitch direction [$\chi^2_{(1)} = 21.098, p < 0.001$], and

music experience and duration [$\chi^2_{(1)} = 4.079, p = 0.043$]. *Post-hoc* tests showed the following: (1) the musicians tended to regard the low vowel [a] stimuli as being the same compared to the high vowel [i], while the non-musicians had the reversed pattern; (2) the musicians were more likely to regard the stimuli as different than the non-musicians, regardless of duration or pitch direction; (3) the low vowel [a] stimuli were more likely to be regarded as the same compared to the high vowel [i] stimuli, regardless of the pitch direction or duration; (4) the rising pitch direction was more likely to be treated as the same compared to the falling pitch direction, regardless of duration values.

Between-Category Discrimination

For each subject, the *d*-prime scores of the comparison unit corresponding to the category boundary were calculated. Linear mixed-effects models were then fitted with *d*-prime scores as the response variable, musical experience, pitch direction, vowel quality, and duration as factors, and subjects as a random effect. The main effects of musical experience [$\chi^2_{(1)} = 7.486, p = 0.006$], pitch direction [$\chi^2_{(1)} = 37.065, p < 0.001$], and duration [$\chi^2_{(1)} = 33.938, p < 0.001$] all reached significance. There were also significant two-way interactions between musical experience and pitch direction [$\chi^2_{(1)} = 29.447, p < 0.001$], musical experience and vowel quality [$\chi^2_{(1)} = 6.495, p = 0.039$], pitch direction and vowel quality [$\chi^2_{(1)} = 6.800, p = 0.033$], and duration and vowel quality [$\chi^2_{(1)} = 7.962, p = 0.004$]. *Post-hoc* tests suggested that musicians had higher *d*-prime scores than non-musicians regardless of pitch direction, duration, or vowel quality. Also, the falling pitch direction had higher *d*-prime scores than the rising pitch direction for musicians, while the reversed pattern was observed for non-musicians. In general, *d*-prime scores improved with the increase of duration for both musicians and non-musicians.

Next, we fitted regression models for Cantonese musicians and non-musicians to capture the relationship between between-category *d*-prime scores (dependent variable) and duration. The model for the musicians (Equation 12) was non-significant [$F_{(1,25)} = 4.412, p = 0.459$; adjusted $R^2 = 0.116$], and the one for non-musicians (Equation 13) was significant [$F_{(1,18)} = 8.014, p = 0.011$; adjusted $R^2 = 0.270$].

$$bcd = 22.710 * d - 3.073 \quad (12)$$

$$bcd = 17.616 * d - 1.975 \quad (13)$$

Within-Category Discrimination

Within-category discrimination *d*-prime scores were calculated for each subject. A linear mixed effects model was fitted where the dependent variable was *d*-prime scores, and the independent variables included musicianship, vowel quality, duration, and pitch direction. The results showed significant main effects of pitch direction [$\chi^2_{(1)} = 140.99, p < 0.001$; mean *d'* scores = 1.441 and 0.777 for falling and rising pitch directions, respectively], and a marginally significant main effect of stimulus duration [$\chi^2_{(1)} = 3.674, p = 0.055$; mean *d'* scores ranged from 0.988 to 1.240]. There was no significant effect of musical experience on the *d*-prime scores.

We then fitted regression models for Cantonese musicians and non-musicians to capture the relationship between within-category d -prime scores (dependent variable) and duration (independent variable). The regression model for the musicians as in Equation (14) was significant [$F_{(1,25)} = 6.747, p = 0.015$; adjusted $R^2 = 0.181$]. The model for non-musicians as in Equation (15) did not reach significance [$F_{(1,18)} = 2.67, p = 0.120$; adjusted $R^2 = 0.081$]. In general, the musicians had higher scores than the non-musicians, although not all pairs showed significant differences.

$$wcd = 0.812 + 5.886 * d \quad (14)$$

$$wcd = 0.305 + 5.775 * d \quad (15)$$

Peakedness

Peakedness of discrimination function was estimated by calculating the difference between P_{bc} (between-category discrimination) and P_{wc} (within-category discrimination). The dependent variable is the peakedness value, and the independent variables included musical experience, pitch direction and stimulus duration. The main effects of musical experience [$\chi^2_{(1)} = 5.828, p = 0.016$] and pitch direction [$\chi^2_{(1)} = 10.434, p = 0.001$] and stimulus duration [$\chi^2_{(1)} = 12.412, p < 0.001$] reached significance. The two-way interactions between musical experience and pitch direction [$\chi^2_{(1)} = 15.469, p < 0.001$] and between pitch direction and duration [$\chi^2_{(1)} = 3.970, p = 0.046$] were also significant. In general, the musicians had larger peakedness than the non-musicians regardless of vowel quality, or duration. The rising pitch direction was always greater in peakedness than the falling pitch direction. There was also a general trend of greater peakedness with the increase of duration for both groups of participants. Moreover, although musicians had larger peakedness than non-musicians for the falling pitch direction, the two groups had comparable peakedness for the rising pitch direction.

Since musical experience was shown to influence the peakedness, we fitted regression models for musicians and non-musicians separately. However, the models suggested that duration did not significantly contribute to the change in peakedness for either musicians [$F_{(1,18)} = 3.341, p = 0.084$; adjusted $R^2 = 0.110$] or for non-musicians [$F_{(1,6)} = 2.43, p = 0.170$; adjusted $R^2 = 0.170$]. The models are listed in Equations (16) (musicians) and (17) (non-musicians).

$$pk = 0.339 - 6.170 * d \quad (16)$$

$$pk = -0.704 + 6.825 * d \quad (17)$$

Predicted and Obtained Discrimination

The scores of the identification and predicted and obtained discrimination for the eight subgroups are plotted in **Figures 6–8**. Three duration values were selected for illustration: 200, 140, and 80 ms. In general, the musicians showed stronger CP than the non-musicians. For both groups, the perception became more categorical with the increase of the stimulus duration.

To confirm these observations, linear regression models were fitted where correlation values between predicted and obtained

discrimination were the dependent variable, and the independent variables included musical experience, vowel quality, duration, and pitch direction. Significant main effects of musical experience [$t_{(46)} = 2.96, p = 0.006$], pitch direction [$t_{(46)} = 2.11, p = 0.044$], and duration [$t_{(46)} = 2.46, p = 0.013$] as well as significant two-way interactions between musical experience and pitch direction [$t_{(46)} = -3.01, p = 0.005$] and between musical experience and duration [$t_{(46)} = -3.20, p = 0.003$] were obtained. These results indicated that, overall, correlation between predicted and obtained discrimination was stronger for non-musicians than musicians and for the rising pitch direction than the falling pitch direction. In addition, correlation for the rising pitch direction obtained from both musicians and non-musicians was stronger than that obtained for the falling pitch direction obtained from musicians [FM < RM, $\chi^2_{(1)} = 6.04, p = 0.014$; FM < RN, $\chi^2_{(1)} = 5.75, p = 0.016$] but not from non-musicians [FN < RN, $\chi^2_{(1)} = 2.35, p = 0.125$], suggesting that identification better predicted discrimination of the rising pitch contour than the falling pitch contour, particularly among musicians.

Next, we calculated the distances between the predicted and obtained discrimination. A linear regression model suggested a significant two-way interaction between musical experience and vowel quality [$t_{(234)} = -2.30, p = 0.022$] and a three-way interaction among musical experience, duration and vowel quality [$t_{(234)} = -2.09, p = 0.021$]. *Post-hoc* tests showed that the pairs FN vs. RN [FN > RN, $\chi^2_{(1)} = 4.146, p = 0.042$] and MA vs. NI [MA > NI, $\chi^2_{(1)} = 4.404, p = 0.036$] were significant. For the three-way interaction, we only found that the distance got larger with the increase of duration, which was true for both groups and for both vowels.

Summary of the Results

In sum, musicians' perception of pitch direction in a non-native language was generally more categorical than non-musicians. Specifically, they exhibited sharper category boundary and were also more sensitive to between-category differences than non-musicians. In addition, they had greater peakedness of the discrimination functions as well as higher correlation between predicted and obtained discrimination. Surprisingly, however, the relationship between between-category discrimination and duration was significant for non-musicians but not for musicians. In contrast, the relationship between within-category sensitivity and duration was significant for musicians, but not for non-musicians. In general, the musicians had higher scores than the non-musicians. In response to our first research goal, these results strongly suggested that musicians with a tonal language background tended to have stronger CP of pitch in a non-native language.

For our second research goal on examining the effects of stimulus duration, we found that both musicians and non-musicians benefited from increased stimulus duration, but musicians were more sensitive than non-musicians to the changes in stimulus duration reflected by more changes in the sharpness of category and the location of category boundary.

Finally, we explored if musicians were more sensitive to factors that may potentially influence pitch processing such

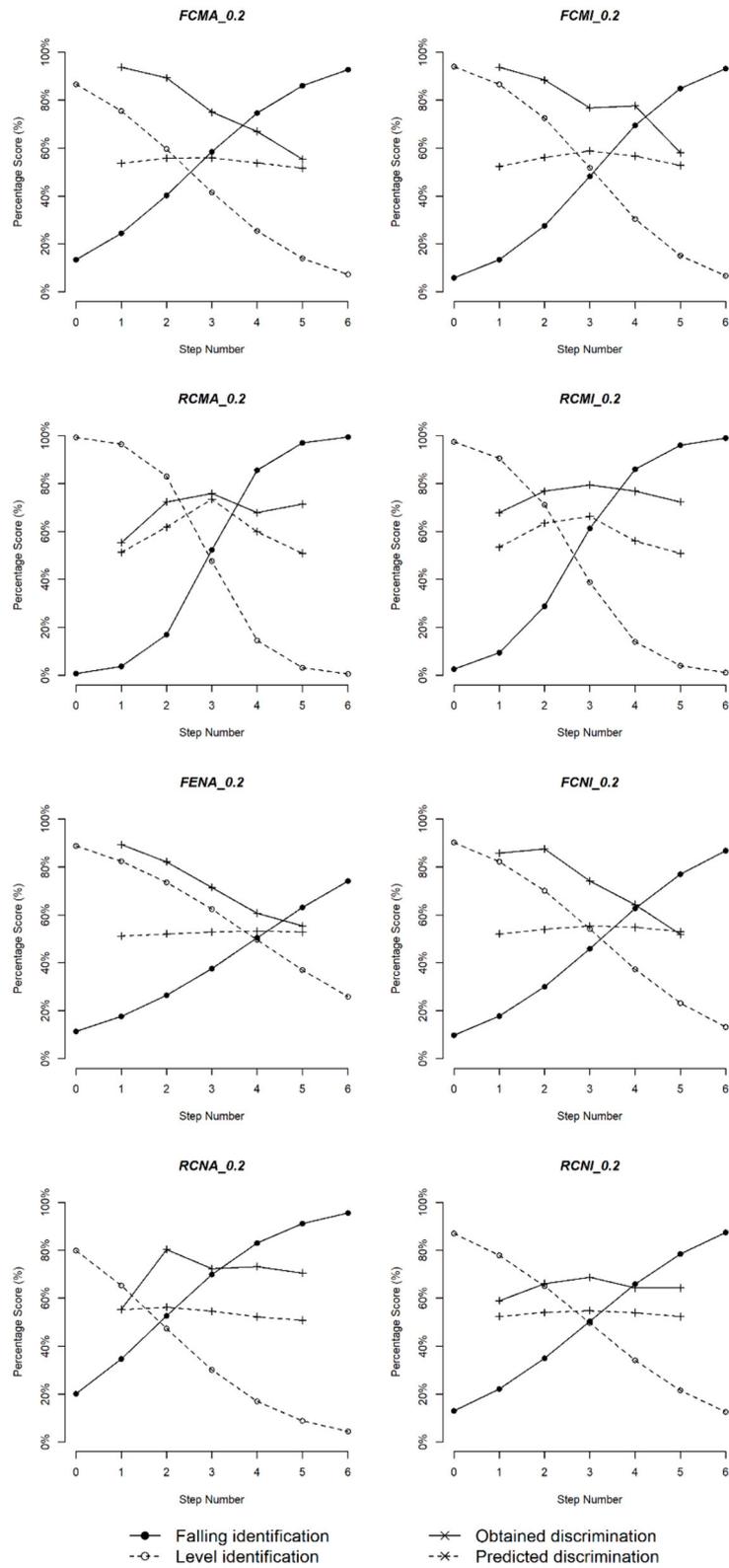


FIGURE 6 | Logistic identification functions and discrimination curves for eight subgroups of Cantonese speakers with duration 200 ms.

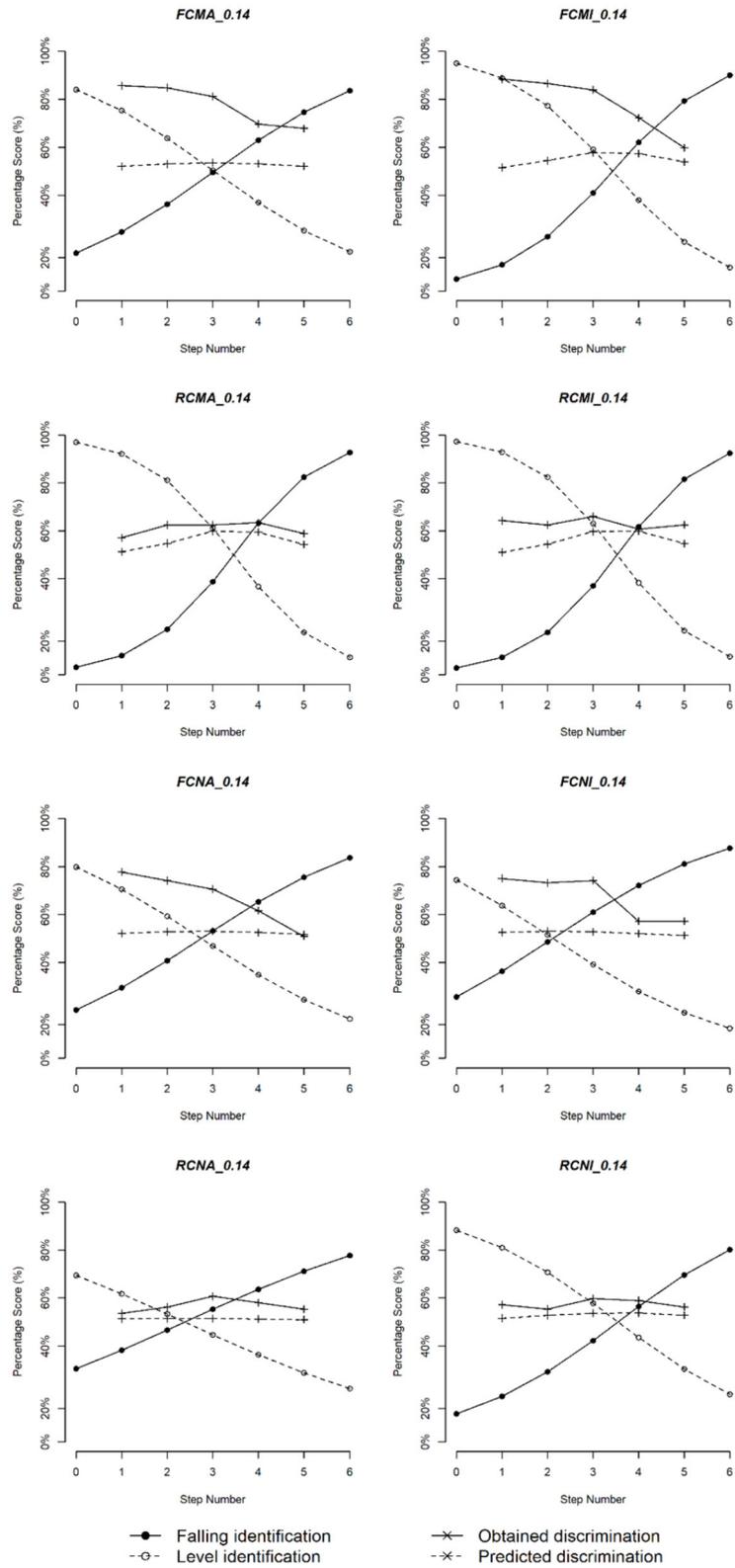


FIGURE 7 | Logistic identification functions and discrimination curves for eight subgroups of Cantonese speakers with duration 140 ms.

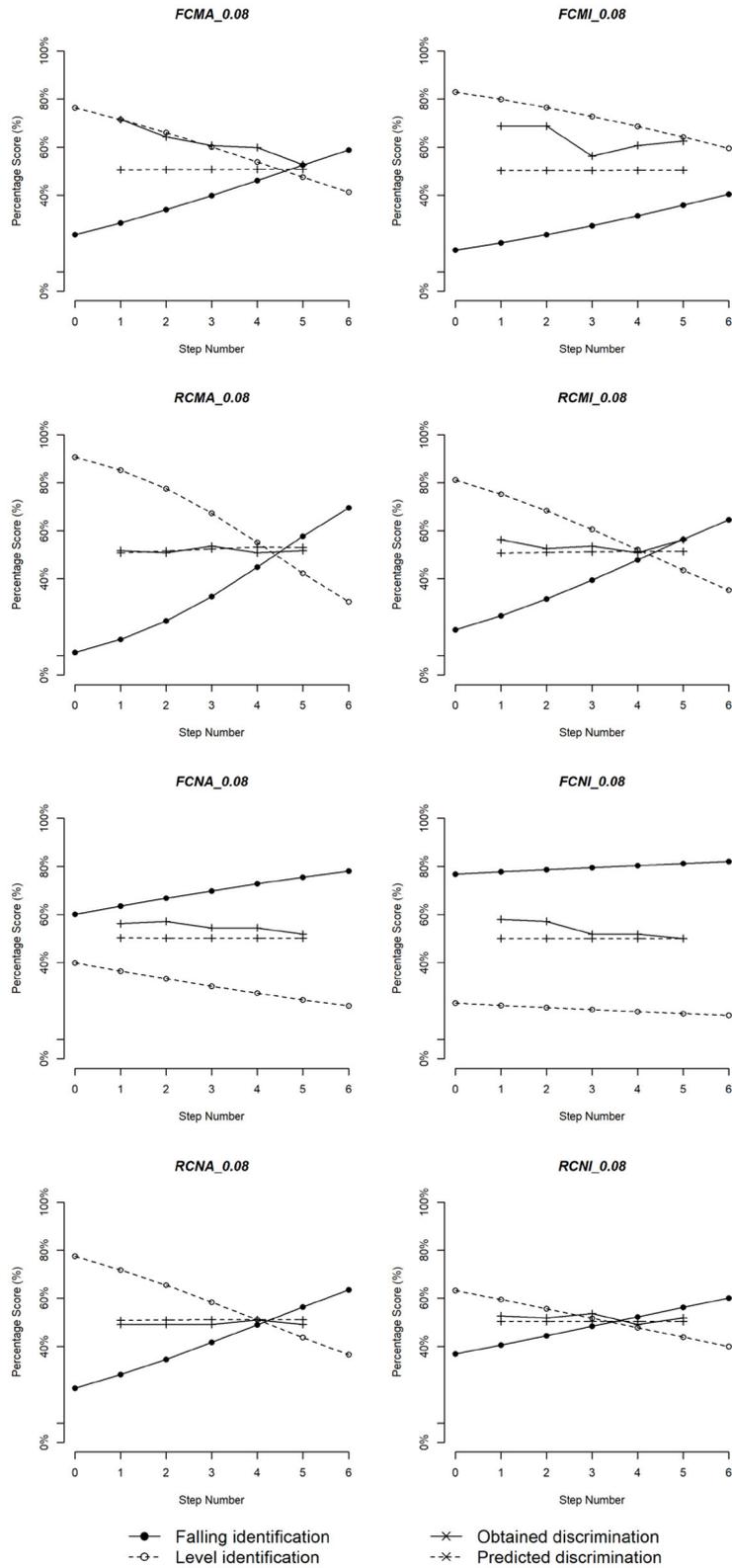


FIGURE 8 | Logistic identification functions and discrimination curves for eight subgroups of Cantonese speakers with duration 80 ms.

as vowel quality and pitch directions. We found that vowel quality did not significantly influence pitch identification among non-musicians, but musicians were more likely to hear more level tones on a high vowel /i/. In addition, both groups showed sharper category boundary for the rising than the falling pitch direction. However, the discrimination results were inconsistent. For between category discrimination, the falling pitch direction had higher d' -prime scores than the rising pitch direction for musicians, but the reversed pattern was observed for non-musicians. For within-category discrimination, the results showed significant higher d' scores for the falling pitch than rising pitch.

DISCUSSION AND CONCLUSION

Using a CP paradigm, this study explored the effects of musical experience, stimulus duration, vowel quality on pitch direction categorization among native Cantonese speakers. Rising and falling pitch direction continua representative of Mandarin Tone 1–2 and Tone 1–4, varying in stimulus duration and vowel height were presented to Cantonese-speaking musicians and non-musicians for identification and discrimination. Sharpness of category derived from the identification data suggested that, in general, musicians exhibited sharper category boundaries than non-musicians, an indication that their perception was more categorical. The results were similar to English-speaking musicians (Chen et al., 2020) and we did not find an obvious ceiling effect of musical experiences reported for the Mandarin-speaking musicians in their study. In addition, albeit statistically non-significance, category boundaries were sharper for the rising pitch direction than the falling pitch direction, and for the high vowel [i] than the low vowel [a] for both groups of listeners. These results suggested the universal psychoacoustic effects of pitch direction and vowel intrinsic F_0 , but a unique effect of musical experience on pitch direction identification. Specifically, rising F_0 and higher intrinsic F_0 associated with a high vowel led to more precise perceptual boundaries between a level and a rising pitch contour among both groups of listeners. However, greater auditory sensitivity to both F_0 parameters (i.e., rising F_0 and higher intrinsic F_0) among musicians than non-musicians resulted in their significantly more abrupt shift in category boundaries.

In addition to sharper identification category boundary, discrimination patterns also suggested that musicians' perception of pitch direction was more categorical than that of non-musicians. Specifically, averaged across vowel height, pitch direction and duration, musicians were more sensitive to between-category differences than non-musicians. Peakedness of the discrimination function also suggested that musicians' perception of pitch direction was more categorical than that of non-musicians. That is, on average, peakedness values were larger for musicians than for the non-musicians, and for the rising than the falling pitch direction. There was also a general trend of greater peakedness with the increase of duration for both groups of participants. Moreover, although musicians had larger peakedness than non-musicians for the falling pitch direction,

the two groups had comparable peakedness for the rising pitch direction. Correlation between predicted and obtained discrimination also pointed to a stronger CP among musicians as it was found that the correlation was stronger for non-musicians than musicians and for the rising pitch direction than the falling pitch direction, particularly among musicians. Regarding distances between predicted and obtained discrimination, they were numerically greater among musicians than non-musicians, but the differences did not reach statistical significance. In addition, the distance became larger as the stimulus duration increases for both groups. The effect of pitch direction was observed only among non-musicians, with greater distance for falling than rising pitch direction. In sum, against characteristics of CP, musicians' perception of pitch direction was generally more categorical than non-musicians. They exhibited sharper category boundary, greater between-category sensitivity, greater peakedness of the discrimination functions, higher correlation between predicted and obtained and greater distance between predicted and obtained discrimination.

In addition to the rising tone being inherently more salient, the possible effects of pitch range cannot be ruled out. As pointed out by an anonymous reviewer, the two Mandarin tonal pairs (T1 vs. T2, T1 vs. T4) are different in the pitch range. According to the traditional five-point scale, the tonal contour changes from point 3 to point 5 for the T2[35]/T1[55] pair, but it changes from point 5 to point 1 for the T1[55]/T4[51] pair. To control for potential confounding effects of differences in acoustic distance between members in both set of stimuli in a CP paradigm, same acoustic differences among stimuli in both tone continua are necessary. However, although the tonal continua are not natural Mandarin tonal continua, a smaller amount of F_0 fall for some members of the falling continuum (e.g., [12], [13]), may render them less "natural" and less acoustically salient than their rising counterparts in the rising continuum. This, in turn, may lead to better discrimination for the rising continuum. However, our discrimination results suggest that this might not always have been the case. For between-category discrimination, the falling pitch direction had higher d' -prime scores than the rising pitch direction for musicians, but the reversed pattern was observed for non-musicians. For within-category discrimination, the results showed significant higher d' scores for the falling pitch than rising pitch. These results suggest that musicality has a stronger effect on discrimination pattern than pitch range.

Our findings suggest that musicians may be at an advantage in learning tones of a second language in that they may better categorize tones in a new tonal language. Cantonese has three level tones and two rising tones, which are different from Mandarin tonal categories (one level tone and one rising tone). Although there might be some positive transfer from Cantonese tones, learners need to establish or revise the existing categories. The beneficial effects from musical training were more obvious for musicians of non-tonal speakers processing tones (Peng et al., 2010; Zhao and Kuhl, 2015a; Chen et al., 2020) than tonal speakers. Our results were inconsistent with what have been found for tone processing in the native language by musicians (Wu et al., 2015; Chen et al., 2020), suggesting that musical training may help tone processing more in a non-native tone

language. It has been argued that musical training help improve auditory memory (Patel, 2011), representations of auditory objects and the mapping between new stimuli and existing memory templates (Bidelman, 2017). In addition, the sensitivity to linguistic and music pitch processing may come from a general cognitive processing (Perrachione et al., 2013). However, musical measures that were indirectly related to tones were less likely to predict tonal word learning (Bowles et al., 2016).

Furthermore, musicians were also more sensitive than non-musicians to changes in stimulus duration when identifying pitch directions. Both musicians and non-musicians benefited from increased stimulus duration, but musicians were more sensitive than non-musicians to the changes in stimulus duration reflected by greater changes in the sharpness of category. In addition, category boundary locations calculated from the identification functions showed that category boundary had a smaller value as the stimulus duration increased for both groups of listeners and musicians had a smaller value than non-musicians. Similar to the findings in Chen et al. (2020), musicians may benefit more from the extra time in that they can use the time to form a more robust auditory representation and matching sounds to internalized memory templates. Since musicians in our study are also native Cantonese speakers, they can rely on existing long-term memory of Cantonese lexical level and rising tonal categories in processing stimuli and further benefit from context-coding, matching sounds to long-term representations.

Both groups were also affected by intrinsic F_0 s associated with vowel height, but only musicians appeared to be perceptually compensated for the effects. Compared to those with lower musical capacity, listeners with higher musical capacity were shown to be more capable of teasing apart acoustic cues and employ more of F_0 , which is a key acoustic cue in pitch processing (Cui and Kuang, 2019). On the other hand, those with lower musical ability tend to rely on both spectral and F_0 cues. Our results thus showed that musical training affords musicians the ability to better accommodate variation in F_0 induced by factors.

While both groups showed similar effects of pitch direction in identification (i.e., sharper category boundary for the rising than the falling pitch direction), the effects of the two pitch directions were inconsistent in discrimination. The identification results were consistent with the reported difficulties that Cantonese speakers have in differentiating Mandarin Tone 1 (high-level) from Mandarin Tone 4 (high-falling) (Hao, 2012). With three contrastive level tones of different F_0 height, Cantonese speakers pay more attention to F_0 height than F_0 contour in comparison to Mandarin speakers (Peng et al., 2012). Also, although Cantonese has two rising tones (Tone 2 [25] and Tone 5 [23]), it only has a falling tone with a shallow slope (Tone 4 [21]), which may lead to their difficulties in identifying falling tones.

Though the factor of proficiency of Mandarin has been controlled between musicians and non-musicians, their exposure to Mandarin may have an effect on CP of their Mandarin lexical tones. Note that non-tonal learners of a tonal language may establish tonal categories and perceive tones in a more categorical way (Shen and Froud, 2016). However, Cantonese learners of Mandarin usually have difficulties in perceiving Mandarin tones. Cantonese speakers pay more attention to F_0

height than F_0 contours compared to Mandarin speakers due to three contrastive Cantonese level tones of different F_0 height in their tonal system (Peng et al., 2012). Also, Cantonese has two rising tones (Tone 2 [25] and Tone 5 [23]), but only a falling tone with a shallow slope (Tone 4 [21]), leading to their difficulties in identifying falling tones. Despite the learning experience of Mandarin, the identification results showed difficulties of Cantonese learners in differentiating Mandarin Tone 1 (high-level) from Mandarin Tone 4 (high-falling) (Hao, 2012). Since Cantonese has both level and rising tones, positive transfer may occur when they perceive Mandarin tones. Our data showed Cantonese musicians had sharper categorical boundary in perceiving rising tones than falling tones, though the differences were less discernible in non-musicians.

Moreover, variability in musical ability and L2 proficiency may influence tone processing. Li and DeKeyser (2017) reported that musical tonal ability is positively correlated with the accuracy rate of perception of tones and comprehension of tone words after training. Also, it has been reported that those with higher musical ability tend to separate acoustic cues better. F_0 is known as a cue most relevant to pitch processing, and listeners with higher musical ability could tease apart spectral cues and F_0 and rely on F_0 only, which may help in tone processing (Cui and Kuang, 2019). Wong and Perrachione (2007) reported that individuals' learning results were positively correlated with their musical experience. Those who received more private music lessons performed better in a Mandarin tone learning task. In addition, it has been reported that prior musical experience and musical aptitude scores predicted success in learning tonal word (Cooper and Wang, 2012) and musical aptitude scores were also positively related to tone discrimination ability (Delogu et al., 2006, 2010). In future studies, it is worth exploring how variability in musical ability and CP are related and whether more musical experience and higher capability will lead to greater sensitivity to vowel quality and more benefits from longer stimulus duration.

The variability in L2 proficiency may also influence tone processing though the results have not been consistent. It has been reported that fluent tone language speakers performed significantly better than less fluent tone language speakers and non-tone language speakers (Deutsch et al., 2009). Hao (2018) tested how proficiency level affects tone discrimination for native English speakers. Three groups of speakers were recruited, including beginning, advanced learners and those who were naïve to Mandarin. Higher accuracy were obtained by learners than those speakers naïve to Mandarin. However, both group of learners were accurate in discriminating tonal pairs except for the T2–T3 pair without significant differences. With extensive exposure to a tone language, adult listeners of non-lexical tone languages may exhibit categorical-like perception for non-native lexical tones. For example, Shen and Froud (2016) reported that native speakers of American English who are advanced learners of Mandarin perceived Mandarin tonal continua in a categorical-like manner, evidencing sharp category boundaries and prominent discrimination peaks. In fact, their discrimination performance was found to be better than that of native Mandarin listeners who had been living as university students in the US for a few years. These results suggest that CP of non-native

phonetic categories can be acquired through intensive and long-term exposure. According to listeners' reports in this study, the participants were proficient non-native Mandarin speakers. It is expected that Cantonese speakers with lower proficiency or naïve to Mandarin may show weaker CP. Therefore, whether musical training may contribute more to tone processing in the initial stage of learning a tone language is worth exploring in future studies.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Human Subject Ethics Sub-committee, the

Hong Kong Polytechnic University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SC and RW contributed to the conception and design of the study. YY collected data performed the statistical analysis based on code written by SC. SC, YY, and RW wrote and revised the manuscript. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

This study was supported by grants from the Department of Chinese and Bilingual Studies [88C3] and [88DW] and Faculty of Humanities [1-ZVHH], [1-ZVHJ], and [ZVNV], the Hong Kong Polytechnic University. We appreciate all the comments from the reviewers and the editor.

REFERENCES

- Abramson, A. S. (1975). "The tones of Central Thai: some perceptual experiments," in *Studies in Thai linguistics in Honor of William J. Gedney*, eds J. G. Harris and J. R. Chamberlain (Bangkok: Central Institute of English Language, Office of State Universities), 1–16.
- Alexander, J. A., Wong, P. C. M., and Bradlow, A. R. (2005). "Lexical tone perception in musicians and non-musicians," in *Proceedings of Ninth European Conference on Speech Communication and Technology* (Barcelona), 397–400. doi: 10.21437/Interspeech.2005-271
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Besson, M., Chobert, J., and Marie, C. (2011). Transfer of training between music and speech: common processing, attention, and memory. *Front. Psychol.* 2, 94. doi: 10.3389/fpsyg.2011.00094
- Bidelman, G. M. (2017). Amplified induced neural oscillatory activity predicts musicians' benefits in categorical speech perception. *Neuroscience* 348, 107–113. doi: 10.1016/j.neuroscience.2017.02.015
- Bidelman, G. M., Gandour, J. T., and Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *J. Cogn. Neurosci.* 23, 425–434. doi: 10.1162/jocn.2009.21362
- Boersma, P., and Weenink, D. (2015). *Praat: Doing Phonetics by Computer*. Available online at: <http://www.praat.org/> (accessed May 01, 2020)
- Bowles, A. R., Chang, C. B., and Karuzis, V. P. (2016). Pitch ability as an aptitude for tone learning. *Lang. Learn.* 66, 774–808. doi: 10.1111/lang.12159
- Burnham, D., Brooker, R., and Reid, A. (2015). The effects of absolute pitch ability and musical training on lexical tone perception. *Psychol. Music* 43, 881–897. doi: 10.1177/0305735614546359
- Chen, S., Zhu, Y., and Wayland, R. (2017). Effects of stimulus duration and vowel quality in cross-linguistic categorical perception of pitch directions. *PLoS ONE* 12, e0180656. doi: 10.1371/journal.pone.0180656
- Chen, S., Zhu, Y., Wayland, R., and Yang, Y. (2020). How musical experience affects tone perception efficiency by musicians of tonal and non-tonal speakers? *PLoS ONE* 15, e0232514. doi: 10.1371/journal.pone.0232514
- Chobert, J., François, C., Velay, J. L., and Besson, M. (2014). Twelve months of active musical training in 8-to 10-year-old children enhances the preattentive processing of syllabic duration and voice onset time. *Cereb. Cortex* 24, 956–967. doi: 10.1093/cercor/bhs377
- Cooper, A., and Wang, Y. (2012). The influence of linguistic and musical experience on Cantonese word learning. *J. Acoust. Soc. Am.* 131, 4756–4769. doi: 10.1121/1.4714355
- Creelman, C. D., and Macmillan, N. A. (2004). *Detection Theory: A User's Guide*. New York, NY: Psychology Press.
- Cui, A., and Kuang, J. (2019). The effects of musicality and language background on cue integration in pitch perception. *J. Acoust. Soc. Amer.* 146, 4086–4096. doi: 10.1121/1.5134442
- Delogu, F., Lampis, G., and Belardinelli, M. O. (2006). Music-to-language transfer effect: may melodic ability improve learning of tonal languages by native nontonal speakers? *Cogn. Process.* 7, 203–207. doi: 10.1007/s10339-006-0146-7
- Delogu, F., Lampis, G., and Belardinelli, M. O. (2010). From melody to lexical tone: musical ability enhances specific aspects of foreign language perception. *Eur. J. Cogn. Psychol.* 22, 46–61. doi: 10.1080/09541440802708136
- Deutsch, D., Dooley, K., Henthorn, T., and Head, B. (2009). Absolute pitch among students in an American music conservatory: association with tone language fluency. *J. Acoust. Soc. Am.* 125, 2398–2403. doi: 10.1121/1.3081389
- Francis, A. L., Ciocca, V., and Ng, B. K. C. (2003). On the (non)categorical perception of lexical tones. *Percept. Psychophys.* 65, 1029–1044. doi: 10.3758/BF03194832
- Fujisaki, H., and Kawashima, T. (1970). Some experiments on speech perception and a model for the perceptual mechanism. *Annu. Rep. Eng. Res. Inst.* 29, 207–214.
- Gottfried, T. L. (2007). "Music and language learning," in *Language Experience in Second Language Speech Learning*, eds O.-S. Bohn and M. J. Munro (Amsterdam: John Benjamins Publishing Company), 221–237. doi: 10.1075/llt.17.21got
- Gottfried, T. L., and Riester, D. (2000). Relation of pitch glide perception and Mandarin tone identification. *J. Acoust. Soc. Am.* 108, 2604. doi: 10.1121/1.4743698
- Gottfried, T. L., Staby, A. M., and Ziemer, C. J. (2004). Musical experience and Mandarin tone discrimination and imitation. *J. Acoust. Soc. Am.* 115, 2545–2545. doi: 10.1121/1.4783674
- Hallé, P. A., Chang, Y. C., and Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *J. Phon.* 32, 395–421. doi: 10.1016/S0095-4470(03)00016-0
- Hao, Y. C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *J. Phon.* 40, 269–279. doi: 10.1016/j.wocn.2011.11.001
- Hao, Y. C. (2018). Second language perception of Mandarin vowels and tones. *Lang. Speech* 61, 135–152.
- Hastie, T. J. and Pregibon, D. (1992) "Generalized linear models," in *Chapter 6 of Statistical Models*, eds J. M. Chambers and T. J. Hastie (Wadsworth & Brooks/Cole).

- Howard, D., Rosen, S., and Broad, V. (1992). Major/minor triad identification and discrimination by musically trained and untrained listeners. *Musical Perception*, 10, 205–220.
- Lee, C.-Y., and Hung, T.-H. (2008). Identification of Mandarin tones by English-speaking musicians and non-musicians. *J. Acoust. Soc. Am.* 124, 3235–3248. doi: 10.1121/1.2990713
- Lee, C.-Y., and Lee, Y.-F. (2010). Perception of musical pitch and lexical tones by Mandarin-speaking musicians. *J. Acoust. Soc. Am.* 127, 481–490. doi: 10.1121/1.3266683
- Lee, Y. S., Vakoch, D. A., and Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: a cross-linguistic comparison. *J. Psycholinguist. Res.* 25, 527–542. doi: 10.1007/BF01758181
- Lehnert-LeHouillier, H. (2013). “From long to short and from short to long: perceptual motivations for changes in vocalic length,” in *Origins of Sound Change: Approaches to Phonologization*, ed A. C. L. Yu (Oxford University Press), 98–111. doi: 10.1093/acprof:oso/9780199573745.003.0004
- Li, M., and DeKeyser, R. (2017). Perception practice, production practice, and musical ability in L2 Mandarin tone-word learning. *Stud. Second Lang. Acquisit.* 39, 593–620. doi: 10.1017/S0272263116000358
- Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *J. Exp. Psychol.* 54, 358–368. doi: 10.1037/h0044417
- Liu, S., and Samuel, A. G. (2004). Perception of Mandarin lexical tones when F_0 information is neutralized. *Lang. Speech* 47, 109–138. doi: 10.1177/00238309040470020101
- Maggu, A. R., Wong, P. C., Antoniou, M., Bones, O., Liu, H., and Wong, F. C. (2018). Effects of combination of linguistic and musical pitch experience on subcortical pitch encoding. *J. Neurolinguistics* 47, 145–155. doi: 10.1016/j.jneuroling.2018.05.003
- Marie, C., Kujala, T., and Besson, M. (2012). Musical and linguistic expertise influence pre-attentive and attentive processing of non-speech sounds. *Cortex* 48, 447–457. doi: 10.1016/j.cortex.2010.11.006
- Mok, P. P., and Zuo, D. (2012). The separation between music and speech: evidence from the perception of Cantonese tones. *J. Acoust. Soc. Am.* 132, 2711–2720. doi: 10.1121/1.4747010
- Moulines, E., and Laroche, J. (1995). Non-parametric techniques for pitch-scale and time-scale modification of speech. *Speech Commun.* 16, 175–205. doi: 10.1016/0167-6393(94)00054-E
- Ong, J. H., Wong, P. C., and Liu, F. (2020). Musicians show enhanced perception, but not production, of native lexical tones. *J. Acoust. Soc. Am.* 148, 3443–3454. doi: 10.1121/1.50002776
- Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Front. Psychol.* 2, 142. doi: 10.3389/fpsyg.2011.00142
- Patel, A. D. (2012). “Language, music and the brain: A resource-sharing framework,” in *Language and Music as Cognitive Systems*, eds P. Rebuschat (Oxford: Oxford University Press).
- Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hear. Res.* 308, 98–108.
- Peng, G., Zhang, C., Zheng, H. Y., Minett, J. W., and Wang, W. S. (2012). The effect of intertalker variations on acoustic-perceptual mapping in Cantonese and Mandarin tone systems. *J. Speech, Lang. Hear. Res.* 55, 579–595.
- Peng, G., Zheng, H. Y., Gong, T., Yang, R. X., Kong, J. P., and Wang, W. S. Y. (2010). The influence of language experience on categorical perception of pitch contours. *J. Phon.* 38, 616–624. doi: 10.1016/j.wocn.2010.09.003
- Perrachione, T. K., Fedorenko, E. G., Vinke, L., Gibson, E., and Dilley, L. C. (2013). Evidence for shared cognitive processing of pitch in music and language. *PLoS ONE* 8, e73372. doi: 10.1371/journal.pone.0073372
- Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Percept. Psychophys.* 13, 253–260. doi: 10.3758/BF03214136
- Pollack, I., and Pisoni, D. (1971). On the comparison between identification and discrimination tests in speech perception. *Psychon. Sci.* 24, 299–300. doi: 10.3758/BF03329012
- Prom-on, S., Xu, Y., and Thipakorn, B. (2009). Modeling tone and intonation in Mandarin and English as a process of target approximation. *J. Acoust. Soc. Am.* 125, 405–424. doi: 10.1121/1.3037222
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: <https://www.r-project.org> (accessed Sep 14, 2020).
- Schneider, W., Eschman, A., and Zuccolotto, A. (2012). *E-Prime User's Guide*. Pittsburgh: Psychological Software Tools Inc.
- Shen, G., and Froud, K. (2016). Categorical perception of lexical tones by English learners of Mandarin Chinese. *J. Acoust. Soc. Am.* 140, 4396–4403. doi: 10.1121/1.4971765
- Stoll, G. (1984). Pitch of vowels: experimental and theoretical investigation of its dependence on vowel quality. *Speech Commun.* 3, 137–147. doi: 10.1016/0167-6393(84)90035-9
- Tang, W., Xiong, W., Zhang, Y. X., Dong, Q., and Nan, Y. (2016). Musical experience facilitates lexical tone processing among Mandarin speakers: behavioral and neural evidence. *Neuropsychologia* 91, 247–253. doi: 10.1016/j.neuropsychologia.2016.08.003
- Wang, W. S.-Y. (1976). “Language change,” in *Origins and Evolution of Language and Speech*, eds S. R. Harnad, H. D. Steklis, and J. Lancaster (New York, NY: New York Academy of Sciences), 61–72. doi: 10.1111/j.1749-6632.1976.tb25472.x
- Wang, W. S. -Y., Lehiste, I., Chuang, C., and Darnovsky, N. (1976). Perception of vowel duration. *J. Acoust. Soc. Am.* 60, S92–S92. doi: 10.1121/1.2003607
- Wang, Y., Yang, X., and Liu, C. (2017). Categorical perception of mandarin chinese tones 1–2 and tones 1–4: effects of aging and signal duration. *J. Speech Lang. Hear. Res.* 60, 3667. doi: 10.1044/2017_JSLHR-H-17-0061
- Whalen, D. H., and Levitt, A. G. (1995). The universality of intrinsic F_0 of vowels. *J. Phon.* 23, 349–366. doi: 10.1016/S0095-4470(95)80165-0
- Wong, P. C., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* 10, 420–422. doi: 10.1038/nrn1872
- Wong, P. C. M., and Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Appl. Psycholinguist.* 28, 565–585. doi: 10.1017/S014271640703312
- Wu, H., Ma, X., Zhang, L., Liu, Y., Zhang, Y., and Shu, H. (2015). Musical experience modulates categorical perception of lexical tones in native Chinese speakers. *Front. Psychol.* 6, 436. doi: 10.3389/fpsyg.2015.00436
- Xu, B. R., and Mok, P. (2012). “Cross-linguistic perception of intonation by Mandarin and Cantonese listeners,” in *Speech Prosody 2012* (Shanghai).
- Xu, Y., Gandour, J. T., and Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *J. Acoust. Soc. Am.* 120, 1063–1074. doi: 10.1121/1.2213572
- Yu, A. C. L., Lee, H., and Lee, J. (2014). “Variability in perceived duration : pitch dynamics and vowel quality,” in *Proceedings of the 4th International Symposium on Tonal Aspects of Languages (TAL 2014)* (Nijmegen), 41–44.
- Zatorre, R. J., and Halpern, A. R. (1979). Identification, discrimination, and selective adaptation of simultaneous musical intervals. *Percept. Psychophys.* 26, 384–395.
- Zhao, T. C., and Kuhl, P. K. (2015a). Effect of musical experience on learning lexical tone categories. *J. Acoust. Soc. Am.* 137, 1452–1463. doi: 10.1121/1.4913457
- Zhao, T. C., and Kuhl, P. K. (2015b). Higher-level linguistic categories dominate lower-level acoustics in lexical tone processing. *J. Acoust. Soc. Am.* 138, EL133–EL137. doi: 10.1121/1.4927632
- Zheng, H. Y., Minett, J. W., Peng, G., and Wang, W. S. (2012). The impact of tone systems on the categorical perception of lexical tones: An event-related potentials study. *Lang. Cogn. Processes* 27, 184–209.
- Zhu, J., Chen, X., and Yang, Y. (2021). Effects of amateur musical experience on categorical perception of lexical tones by native chinese adults: an ERP study

[Internet]. *Front. Psychol.* p. 690. Available online at: <https://www.frontiersin.org/article/10.3389/fpsyg.2021.611189>

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in

this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Chen, Yang and Wayland. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.