The effect of Mandarin listeners' musical and pitch aptitude on perceptual learning of Cantonese level-tones

Zhen Qin, Caicai Zhang and William Shi-yuan Wang

Citation: The Journal of the Acoustical Society of America **149**, 435 (2021); doi: 10.1121/10.0003330
View online: https://doi.org/10.1121/10.0003330
View Table of Contents: https://asa.scitation.org/toc/jas/149/1
Published by the Acoustical Society of America

JASA ARTICLE

The effect of Mandarin listeners' musical and pitch aptitude on perceptual learning of Cantonese level-tones

Zhen Qin,[1,a)] Caicai Zhang,[2,b)] and William Shi-yuan Wang[2]

[1]*Division of Humanities, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong*

[2]*Research Centre for Language, Cognition, and Neuroscience, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Yuk Choi Road, Hung Hom, Hong Kong*

ABSTRACT:

Contrary to studies on speech learning of consonants and vowels, the issue of individual variability is less well understood in the learning of lexical tones. Whereas existing studies have focused on contour-tone learning (Mandarin) by listeners without experience of a tonal language, this study addressed a research gap by investigating the perceptual learning of level-tone contrasts (Cantonese) by learners with experience of a contour-tone system (Mandarin). Critically, we sought to answer the question of how Mandarin listeners' initial perception and learning of Cantonese level-tones are affected by their musical and pitch aptitude. Mandarin-speaking participants completed a pretest, training, and a posttest in the level-tone discrimination and identification (ID) tasks. They were assessed in musical aptitude and speech and nonspeech pitch thresholds before training. The results revealed a significant training effect in the ID task but not in the discrimination task. Importantly, the regression analyses showed an advantage of higher musical and pitch aptitude in perceiving Cantonese level-tone categories. The results explained part of the level-tone learning variability in speakers of a contour-tone system. The finding implies that prior experience of a tonal language does not necessarily override the advantage of listeners' musical and pitch aptitude.

 https://doi.org/10.1121/10.0003330

(Received 3 June 2020; revised 9 December 2020; accepted 16 December 2020; published online 21 January 2021)

[Editor: Benjamin V, Tucker] Pages: 435–446

I. INTRODUCTION

Individual differences are a central issue in research on second language (L2) speech learning, which refers to the observation that some listeners are better than others in learning to discriminate and/or identify non-native speech sound categories (Darcy *et al.*, 2015; Golestani and Zatorre, 2009; Wanrooij *et al.*, 2013). Previous studies have identified various sources of individual differences in L2 perceptual learning of consonants (Bradlow *et al.*, 1997; Flege *et al.*, 1996; Polka, 1991; Schertz *et al.*, 2015) and vowels (Díaz *et al.*, 2008; Escudero *et al.*, 2009; Flege *et al.*, 1997; Kim *et al.*, 2018). L2 learners' acquisition of lexical prosody, such as lexical tones, has been the focus of relatively less research and is accordingly less well understood (Hao, 2012; Leather, 1987; Pelzl, 2019; Sun, 1997; Wagner-Gough and Hatch, 1975). In tonal languages, pitch variations are used to distinguish lexical meanings. For successful perceptual learning of lexical tones, the ability to distinguish and categorize pitch variations is, thus, indispensable (Bent *et al.*, 2006; Ho, 1976; Li and Thompson, 1977; Lin, 1985; Qin *et al.*, 2019; Tse, 1978; Wang *et al.*, 1999). Individual differences in learning Mandarin (contour) tone-word association have been relatively well documented through intensive training of Mandarin tones (Bowles *et al.*, 2016; Dong *et al.*, 2019; Perrachione *et al.*, 2011; Sadakata and McQueen, 2014). Specifically, several studies have demonstrated that English- or Dutch-speaking L2 learners with higher cognitive aptitude (i.e., good learners) exhibited both better learning (i.e., higher perception accuracy in the final training session) and faster learning (i.e., faster improvement of perception accuracy over training sessions) of non-native tones than did their peers with lower aptitude (i.e., poor learners; Bowles *et al.*, 2016; Sadakata and McQueen, 2014). Such individual differences in tone learning outcomes were suggested to be related to the learners' musical and pitch aptitude, which was assessed as their ability to detect a difference in melody or rhythm (e.g., The Advanced Measures of Music Audiation) and identify the (nonspeech) pitch contour (e.g., as "flat," "rising," or "falling" pitch contours), respectively, in Bowles *et al.* (2016). Different from studies examining contour-tone learning by learners without experience of a tonal language (Bowles *et al.*, 2016; Sadakata and McQueen, 2014), the objective of this study is to investigate how musical and pitch aptitude of Mandarin-speaking participants modulates their perception and learning of Cantonese level-tone contrasts. Therefore, the findings of this study may help to explain the source of the less-documented variability of level-tone learning abilities by learners with tonal language experience.

A. The role of musical aptitude in tone perception and learning

A first line of evidence indicates an advantage of higher musical aptitude in tone learning for nontonal language

[a)]ORCID: 0000-0002-6117-7809.

[b)]Electronic mail: caicai.zhang@polyu.edu.hk, ORCID: 0000-0002-7687-0518.

speakers with musical experience (Lee *et al.*, 2014; Wong *et al.*, 2007; Wong and Perrachione, 2007; Zhao and Kuhl, 2015). Music and speech share many sound attributes, including the systematic use of pitch (Jackendoff and Lerdahl, 2006; Patel, 2008). While pitch differences are used to form musical melodies, they are used contrastively in speech at the utterance level (intonation) and word level (lexical tones; Ladd, 1996). Given the shared use of pitch, pitch patterns in one domain (music) can influence pitch processing in another domain (speech) and vice versa (Patel, 2008). That is, there could be a cross-domain transfer of musical aptitude and language experience on pitch processing (Chen *et al.*, 2016; Zhang, 2019). Previous research provided substantial evidence supporting the close relationship between musical experience and lexical tone perception by investigating the effect of musicianship on tone perception by speakers of nontonal languages (e.g., English). For instance, these studies have found that English-speaking musicians with higher musical aptitude outperformed English-speaking nonmusicians in identifying (Lee *et al.*, 2014; Lee and Hung, 2008) and discriminating Mandarin tones (e.g., Alexander *et al.*, 2005).

Likewise, the effect of musical aptitude on tone perception was also found among nonmusicians without prior experience of a tonal language who exhibited a spectrum of individual variations of musical aptitude (Cui and Kuang, 2019; Li and DeKeyser, 2017; Zhang *et al.*, 2018). Li and Dekeyser (2017) measured musical aptitude of English-speaking nonmusicians using a pitch change test and a perceived musical memory test selected from the Wing measures of musical talents (WMMT; Farnsworth, 1969). The English-speaking participants did not receive more than three years of musical training in any musical instrument (including vocal singing). They were instructed to learn Mandarin tone-word mapping via either a word comprehension task or a word production task during a short-term training session. The results indicated that the English-speaking learners' musical aptitude was a significant predictor for their tone identification (ID) accuracy but not their tone production accuracy in the posttests. The finding suggested that the nonmusicians' musical aptitude may also predict their perceptual learning outcome of non-native tones.

However, it remains less clear how musical aptitude affects tone perception, especially *non-native* tone perception, in tonal language speakers. A few studies have examined the effect of musical aptitude on *native* tone perception. However, the effect was not always detected and might be sensitive to the specific task or measure employed (Lee *et al.*, 2011; Lee and Lee, 2010; Sadakata *et al.*, 2020). Several studies did not find a difference between musicians and nonmusicians of tonal language speakers in their perception of native tones (Mok and Zuo, 2012; Tong *et al.*, 2018). On the other hand, a few studies reported that Mandarin-speaking musicians showed an advantage over their nonmusician peers when discriminating within-category pitch distinctions (Wu *et al.*, 2015) or in the response speed of Mandarin tone perception (Tang *et al.*, 2016). It may be argued that because of the strong influence of top-down language knowledge, the effect of musical aptitude may be restricted in native tone perception. Indeed, it has been found that native Mandarin listeners are capable of achieving fairly accurate native tone perception (i.e., above-chance performance) even when pitch information in the speech stimuli is absent (e.g., whispered), neutralized (e.g., pitch-flattened speech), or partially removed (e.g., brief acoustic input with only the first six glottal periods remained; Jiao and Xu, 2019; Lee, 2009; Xu *et al.*, 2013). However, in the case of *non-native* tone perception in tonal language speakers, a different scenario may be present. In a recent study, Cui and Kuang (2019) used the Montreal Battery of Evaluation of Amusia (MBEA) to assess listeners' musical aptitude, that is, the ability to detect differences in melody, rhythm, as well as the memory of musical phrases (Peretz *et al.*, 2003; Peretz *et al.*, 2013). The authors tested the effect of musical aptitude on the use of pitch cues and spectral cues in a pitch classification task. The results showed an effect of musical aptitude on the listeners' relative weighting of different cues with a higher musical aptitude predicting a greater use of pitch relative to spectral cues in judging specific pitch patterns. Crucially, this effect was found for both Mandarin and English listeners regardless of whether they speak a tonal language or not. However, it should be noted that the authors employed a general-purpose pitch classification task (i.e., comparing the pitch height of two meaningless sequences ma-MA-ma and ma-MA-ma with various pitch and spectral manipulations). The question of how the musical aptitude of tonal language speakers affects the perception of non-native lexical tones remains unexplored.

Taken together, previous research tested how musical aptitude influenced native tone perception and general pitch perception by tone-language speakers. However, little is currently known regarding the effect of musical aptitude on non-native tone perception and learning in speakers of a tonal language. This study undertakes this investigation and tests how the musical aptitude of Mandarin-speaking learners predicts their perception and learning of non-native Cantonese tones.

B. The use of pitch height in tone perception and learning

In addition to the role of musical aptitude, pitch processing ability assessed as sensitivity to pitch contour (i.e., falling vs rising contour) has also been associated with a better outcome for Mandarin contour-tone word learning (Bowles *et al.*, 2016; Chandrasekaran *et al.*, 2010; Gandour, 1983; Ingvalson *et al.*, 2013; Perrachione *et al.*, 2011; Sadakata and McQueen, 2014). For instance, Perrachione *et al.* (2011) used a Pitch-Contour Perception Test (PCPT) to assess English-speaking participants' pitch contour processing abilities prior to the training of Mandarin tones. During the test, the participants identified the pitch contours (level, rising, and falling) that they heard by matching the

https://doi.org/10.1121/10.0003330

auditory stimuli to representative arrows that mimic the pitch direction on the computer screen. After a word-tone association training, the learners' PCPT scores were found to be predictive of word-tone learning outcomes regardless of the training paradigm and learners' individual language aptitude. Whereas pitch contour distinguishes different contour tones in Mandarin and has been tested in quite a few studies, pitch height (i.e., average height; higher vs lower tones) is an under-investigated dimension of tone perception and learning. It should be noted that pitch height also plays an important role in contrasting level-tones in other tonal languages (Gandour, 1981, 1983). For instance, in the Cantonese tonal inventory {/si 55/, "silk" [tone 1 (T1)]; /si 25/, "history" (T2); /si 33/, "to try" (T3); /si 21/, "time" (T4); /si 23/, "city" (T5); and /si 22/, "matter" (T6)}, level-level-tone pairs such as T3 (mid-level)–T6 (low-level) are distinguished by subtle differences in the pitch height (Mok and Zuo, 2012). Such tones with less dynamic contour change and, yet, fine-grained pitch height differences could be perceptually challenging even for listeners with tonal language experience (Francis and Ciocca, 2003). In line with this idea, it has been found that Cantonese level-level tonal contrasts posed a great perceptual difficulty for Mandarin listeners (Chang *et al.*, 2017; Jongman *et al.*, 2017). A previous study tested the perception of Cantonese contour-level and level-level tonal contrasts by the Mandarin-speaking and English-speaking participants in an AX (same-different) discrimination task (Qin and Jongman, 2016). The results of *d*-prime scores (computed for each participant based on a difference between the "hit" rate and the "false alarm" rate, serving to tease apart any response bias from sensitivity) showed that the English listeners were better than the Mandarin listeners in using pitch height to discriminate the level-level-tone pairs. The Mandarin listeners' reduced sensitivity to pitch height was attributed to their native tonal inventory in which no tones are contrastive in pitch height alone. It is likely that the fine-grained differences in pitch height among Cantonese level-tones were treated as within-category differences by Mandarin listeners.

To summarize, since the sensitivity to pitch height is crucial for perceiving level-level tonal contrasts, it remains unclear whether learners' sensitivity to pitch height would predict their ability to learn Cantonese level-level tonal contrasts (Qin and Zhang, 2019). This issue is addressed in the current study.

C. The current study

Little research (to our knowledge) to date has investigated individual differences of learning non-native level-tone contrasts in tonal language speakers. To this end, the present study examines whether, and if so how, Mandarin listeners' perceptual learning (as well as initial perception before training) of Cantonese level-level tonal contrasts is predicted by their musical aptitude and pitch height processing abilities.

With regard to musical aptitude, the MBEA has been used to quantify listeners' musical aptitude as continuous musicality scores in previous studies (Chen *et al.*, 2016; Cui and Kuang, 2019). Thus, this study used the MBEA to measure musical aptitude as continuous musicality scores.

Regarding the pitch height processing ability, the psychoacoustic pitch threshold has been used to assess listeners' auditory processing of pitch and proven to account for individual differences of L2 speech (i.e., vowels) learning outcomes (Kachlicka *et al.*, 2019). Therefore, following the designs of Kachlicka *et al.* (2019) and Liu *et al.* (2012), this study measured listeners' thresholds for identifying pitch height using discrete pitch stimuli in both the speech and nonspeech domains (Bowles *et al.*, 2016).

Based on the finding that musical and pitch aptitude modulates tone learning by nontonal language speakers (Li and DeKeyser, 2017; Perrachione *et al.*, 2011) and the evidence of musical advantage in tone perception by tonal language speakers (Cui and Kuang, 2019; Tang *et al.*, 2016), it is hypothesized that Mandarin listeners' musical aptitude and pitch height processing abilities will predict their perceptual learning (and initial perception before training) of Cantonese level-tones.

II. METHODS

A. Participants

Thirty-two students (18 female, 14 male, aged 18–30 years old) were recruited from the Hong Kong Polytechnic University. They were all native Mandarin speakers with a minimal exposure to Cantonese; that is, their length of residence in Hong Kong was shorter than ten months and classroom learning of Cantonese was less than one month prior to the experiment. They speak Mandarin as their mother tongue and not any Southern Chinese dialect (e.g., Shanghainese). Following the criterion of nonmusicians in Li and Dekeyser (2017), we only recruited participants who had not received more than three years of music lessons in any musical instrument, including vocal training. None of them reported a history of hearing impairment and neurological disorders.

The experimental procedures were approved by the Human Subjects Ethics Subcommittee of the Hong Kong Polytechnic University. Informed written consent was obtained from the participants in compliance with the experiment protocols. All of the participants were paid for their participation.

B. Stimuli

The stimuli were three Cantonese level-tones, T1 (/55/ a high-level-tone), T3 (/33/, a mid-level-tone), and T6 (/22/, a low-level-tone) carried by ten base syllables (/jan/, /ji/, /jau/, /jiu/, /fan/, /fu/, /ngaa/, /si/, /se/, and /wai/). All 30 words are meaningful in Cantonese (and their counterparts, most of which have different pronunciations, can be found in Mandarin).

Two female speakers of Hong Kong Cantonese with different pitch ranges recorded three repetitions of each target word. Each monosyllabic target word was recorded in a carrier phrase "lei1 go3 hai6 [target word]" (this is [target word]). One of the two speakers was used in the training, and the other speaker was used in the posttest to assess generalization to a novel talker (see details in Sec. II C). Recording was conducted in a sound attenuated booth using Praat on a personal computer (PC) workstation connected with an Azden ECZ990 microphone (Mt. Arlington, NJ). The recordings were made at a sampling rate of 44 000 Hz with 16 bits per sample. Each token was segmented out of the carrier phrase by Z.Q. To increase the variability of tone stimuli, two tokens for each target word were chosen from the three repetitions by the investigators based on their intelligibility and pronunciation accuracy. The stimuli were normalized in duration to 500 ms (a value similar to the duration of naturally produced stimuli), and their mean acoustic intensity was scaled to 70 dB using Praat (Boersma and Weenink, 2018).

C. Procedure

The study adopted a pretest-posttest design for tone training (Earle and Arthur, 2017; Wang *et al.*, 1999; Wayland and Li, 2008). Participants completed a pretest, training, and a posttest over two days within one week. On the first day (day 1), participants were tested in a set of pretests which measured their musical aptitude and pitch-height threshold. On the second day (day 2), participants began the session with an AX discrimination pretest, followed by a training session of tone ID, and finished with a posttest session (see detailed descriptions below).

1. Pretests: Musical perception and pitch threshold tasks

First, the participants' musical aptitude was assessed using the MBEA (Peretz *et al.*, 2003). To test the participants' musical aptitude comprehensively, following Cui and Kuang (2019), all of the six subtests of MBEA, which focus on different aspects of musical perception, were included. The first three subtests are pitch-related tasks that concern melodic organization (scale, the pitch related to tonal functions; contour, direction of pitch sequences in a melody; interval, pitch range size between two different melodies). The next two subtests are duration-related tasks that concern temporal organization (rhythm and meter). The last subtest is a memory task that concerns the incidental memory of melodies from the preceding pitch-related and duration-related subtests.

In the three pitch-related tasks and the rhythm task, listeners listened to a pair of short melodies in the piano timbre in each trial and were required to judge whether each pair was identical or different. Half of the items were different pairs in which the musical scale/contour/interval or the rhythm was altered in the second melody of a pair, whereas the other half had identical melodies or rhythms. In the meter task, listeners listened to a melody in each trial and were required to categorize each melody as a waltz or a march with regard to the temporal regularity or pulse of a melody. The number of waltz and march melodies was counterbalanced in this task. The memory task was used to assess the listeners' incidental musical memory, following *implicit* storage of melodies presented in the previous MBEA subtests because the participants were not informed in advance that their memorization of melodies would be tested. In this task, listeners listened to 15 previously presented melodies and 15 new melodies, which were pseudorandomly mixed. The task was to identify whether they had previously heard the melody or not. Each subtest of the MBEA consisted of 30 experimental trials, and there were 180 trials in total. For each participant, a composite score of MBEA-pitch subtests [mean = 0.82; standard deviation (SD) = 0.09; range, 0.60–0.97] was calculated by using the listeners' proportion of correct responses averaged across the three pitch-related subtests; a composite score of MBEA-duration subtests (mean = 0.79; SD = 0.13; range, 0.40–0.98) was calculated by averaging the listeners' proportion of correct responses in the rhythm and meter subtests; the MBEA musical memory score (mean = 0.91; SD = 0.10; range, 0.58–1.0) was the listeners' proportion of correct responses in the memory subtest. The composite scores of different MBEA subtests, instead of an overall musicality score, were calculated to test the specific dimensions of musical aptitude (pitch, duration, and memory).

Second, the participants' sensitivities of pitch height were assessed in a pitch threshold test. The test used in the study was identical to that reported in Ho *et al.* (2018) and follows the design of Liu *et al.* (2012). The speech and nonspeech stimuli were included to test listeners' pitch thresholds in both speech and nonspeech domains in a comprehensive manner (Bowles *et al.*, 2016). Regarding the speech tones, participants were assessed on speech tones that were carried by the Cantonese syllable /ji/, produced by a male native speaker of Cantonese. Regarding the nonspeech tones, the complex tones (not pure tones), which had more than a single frequency component and carried the same F0 as the speech tones, were generated using Praat (Boersma and Weenink, 2018). A 15-ms amplitude ramp was applied to the onset and offset of the complex tones to adjust for rise or decay time. All stimuli had duration normalized to 250 ms with a 250 ms interstimulus interval. The task was a two-alternative forced choice in which listeners were instructed to click a mouse button to indicate the pitch pattern of each stimulus pair as high-low or low-high. Each trial included a standard stimulus of 100 Hz paired with a stimulus of 82 target stimuli ranging from 100.07 to 178.17 Hz in steps of 0.01, 0.1 and 1 semitone. The trials began with a 10 semitone difference and followed a "two-down, one-up" staircase method design. The semitone difference within a pair was reduced by 1 semitone upon two consecutive correctly judged trials, and the reduction was adjusted to 0.1 semitone when the semitone difference reached 1 semitone and adjusted to 0.01 semitone when the semitone difference reached 0.1 semitone; an incorrect trial led to a reversal to the previous semitone difference. The task ended after 14 reversals, and the pitch threshold (speech, mean = 1.7; SD = 2.6; range, 0.2–8.8; nonspeech, mean = 2.2; SD = 3.3; range, 0.2–9.6) was calculated as the mean pitch difference in semitones between the standard and target stimuli in the last six reversals. The order of speech and nonspeech conditions was counterbalanced across the participants.

The MBEA test and pitch threshold task were conducted using E-Prime, version 2.0 (Psychology Software Tools, Pittsburgh, PA). All stimuli were presented in a sound attenuated room through JVC HA-D610 stereo headphones (JVC, Yokohama, Japan) binaurally at a comfortable listening level, which was chosen by the participants at the beginning and kept constant throughout the experiment.

2. Main tests: Training, AX discrimination tests, and ID posttests

An AX discrimination task (details provided in the posttest session), which does not require a tone-category pairing, was conducted to measure listeners' initial tone perception abilities immediately before the training session.

In the training session, a forced-choice ID task of the three Cantonese level-tone categories was administered to instruct the participants' learning of the tone-category pairing. To control the learning difficulty of the training stimuli, only tone stimuli produced by the first female speaker, who had a wider range of pitch than the other speaker, were used in the training. A total of 300 tokens (1 speaker × 3 tones × 10 syllables × 2 tokens × 5 repetitions) were presented in an auditory mode to the participants with 60 tokens in a block, and each block was repeated 5 times. During the training, the participants were instructed to identify each tone (T1-high, T3-mid and T6-low) after hearing the auditory stimuli by pressing three buttons (1, 3, and 6) in a self-paced fashion. Written feedback ("Correct" in green or "Incorrect. The correct answer is…" in red) in English was given immediately after every trial. The participants were instructed to learn to categorize three tones based on feedback and achieve the best performance that they can in this session. The training session started with a block consisting ten practice trials (different from the experimental trials).

The posttest session was conducted immediately after training to examine the learning outcome of the three Cantonese level-tones. To test whether the participants learned to categorize novel tones at a phonological level instead of an acoustic level, tone stimuli produced by both female speakers were used in this session. Both ID and AX discrimination tasks were used to assess the learning outcome. In the ID posttest, similar to the training session, the participants were instructed to identify each tone (T1-high, T3-mid and T6-low), with 3 s as the time-out limit, by pressing the three buttons (1, 3, and 6). However, no feedback was given after the response in each trial. A total of 120 tokens (2 speakers × 3 tones × 10 syllables × 2 tokens) were randomly presented to participants in 1 block. The ID posttest took approximately 5 min.

An AX discrimination posttest, which was identical to the AX discrimination pretest and considerably longer than the ID posttest, was administered after the ID posttest to reduce any potential effect of fatigue after a long task. In the AX discrimination task that was used in both the pretest and posttest sessions, the participants were instructed to distinguish whether the two tones they heard belonged to the same or different tone categories by pressing one of the two buttons (left arrow and right arrow), indicating "the same" or "different," respectively, on the keyboard. The two stimuli within each pair were always carried by the same syllable and only contrasted in tones. The interstimulus interval was 1000 ms. No feedback was given. An equal number of AA pairs (120 pairs with the same tone within each pair) and AB pairs (120 pairs with different tones within each pair) were used to counterbalance the two types of tone pairs. The presentation order of two tones in each AB pair was also counterbalanced in different trials. Two acoustically different tokens of the same tone type were used in each AA pair such that the listeners would need to discriminate tones based on whether there was a change in the identity of tone categories.

A total of 240 tokens (2 speakers × 3 pairs × 2 orders × 10 syllables × 2 types) were presented in a random order to the participants in 1 block. The AX discrimination posttest took approximately 20 min. The AX discrimination and ID tasks were conducted using the Paradigm software (Perception Research Systems, Inc., available online[1]).

D. Data analyses

A series of multiple regression analyses were conducted to address whether, and if so how, different aspects of musical aptitude, as well as pitch threshold of speech and nonspeech tones, predict the initial perception and learning of novel tone categories.

In the regression analyses, the composite scores of MBEA-pitch subtests, MBEA-duration subtests, and MBEA musical memory scores were treated as continuous variables and entered as predictors. The speech pitch threshold and the nonspeech pitch threshold in semitones were log-transformed to adjust a highly positive skew of the raw data, and then were entered as predictors (Kachlicka *et al.*, 2019; Larson-Hall, 2015). All the predictor variables were z-normalized and centered prior to the analyses. Following previous studies (Earle *et al.*, 2017; Earle and Arthur, 2017), these variables were entered stepwise as predictors into a multiple regression model using IBM SPSS Statistics for Windows, version 25 (IBM Corp., Armonk, NY). Only the predictors that contributed to improving the analysis remained in the model.

The dependent variables included the participants' performance in the AX discrimination pretest, AX discrimination posttest, and ID posttest. For the AX discrimination pretest and posttest, the participants' responses were

https://doi.org/10.1121/10.0003330

TABLE I. The regression model on *d*-prime scores of the AX discrimination pretest. *β*, standardized coefficients of beta; SE, coefficients standard error.

Variables	Predictors	*β*	SE	*t*	*p*
d-prime scores of AX discrimination pretest	Intercept		0.07	34.55	<0.001
	MBEA musical memory	0.60	0.07	4.56	<0.001
	Pitch threshold of nonspeech tones	−0.36	0.12	−2.71	0.01

converted to *d*-prime (pretest, mean = 2.62; SD = 0.55; posttest, mean = 2.61; SD = 0.48), which was commonly applied to speech discrimination tasks (Francis and Ciocca, 2003; Macmillan and Creelman, 2004). Specifically, the *d*-prime score for each participant was computed based on the hit rate of his/her AB trials (number of times the "different" button pressed for AB pairs) and the false alarm rate of his/her AA trials (number of times the different button pressed for AA pairs). A *d*-prime score is the difference between the hit rate and false alarm rate after they are *z*-normalized. For the ID posttest, the dependent variable was the participants' proportion of correct responses (i.e., accuracy; mean = 0.64; SD = 0.12).

III. RESULTS

A. Initial tone perception

To test whether listeners' musical and pitch aptitude predicts their initial ability to discriminate novel tones, first, we regressed *d*-prime scores of the AX discrimination pretest with MBEA-pitch, MBEA-duration, MBEA musical memory, speech pitch threshold, and nonspeech pitch threshold as predictors. The final model significantly accounted for variance in the *d*-prime scores of the AX discrimination pretest ($F_{2,29} = 14.49$, $p < 0.001$; adjusted $R^2 = 0.47$). A *post hoc* analysis was conducted using G*Power 3.1 (Erdfelder *et al.*, 2009) to test the power of the regression model with an alpha of 0.05. The power of the model achieved was 0.98. Moreover, the variance inflation factors of the predictors remaining in the model were checked. The values of the predictors were between 1 and 1.5, which indicated a low multicollinearity (Larson-Hall, 2015).

Table I summarizes the regression model on *d*-prime scores of the AX discrimination pretest with musical memory and nonspeech pitch threshold as significant predictors. In the model, the effect of the musical memory and nonspeech pitch threshold independently accounted for a significant portion of the variance after adjusting for excluded predictors, MBEA-pitch scores ($\beta = -0.20$, $t = -1.09$, $p = 0.28$), MBEA-duration scores ($\beta = -0.08$, $t = -0.49$, $p = 0.63$), and speech pitch threshold ($\beta = -0.07$, $t = -0.37$, $p = 0.71$).

Figure 1 shows how the listeners' *d*-prime scores of the AX discrimination pretest were predicted by their musical memory and nonspeech pitch threshold. Specifically, a higher *d*-prime score of the AX discrimination pretest was associated with a higher musical memory score and a lower (more sensitive) nonspeech pitch threshold. There was no significant relationship between *d*-prime scores of the AX discrimination pretest and other predictors. The results suggested that the Mandarin listeners' incidental musical memory abilities and their pitch thresholds of nonspeech tones (but not that of speech tones) predicted their initial discrimination abilities of Cantonese level-tones.

B. Tone learning: Tone discrimination

To test whether the listeners showed a learning effect in the AX discrimination task, a paired-samples *t*-test was conducted on the listeners' *d*-prime scores of their AX discrimination pretest (mean = 2.62; SD = 0.56) vs posttest (mean = 2.61; SD = 0.49). There was no significant difference [$t(31) = 0.25$, $p = 0.80$] between the pretest and posttest.

The listeners' posttest discrimination performance could be contingent on their pretest discrimination performance. To determine whether the relationships between the listeners' posttest discrimination abilities and the musical and pitch predictors were independent of or epiphenomenal to the listeners' pretest discrimination abilities, we, therefore, ran the same regression analysis on the *d*-prime scores of the AX discrimination posttest as above but additionally

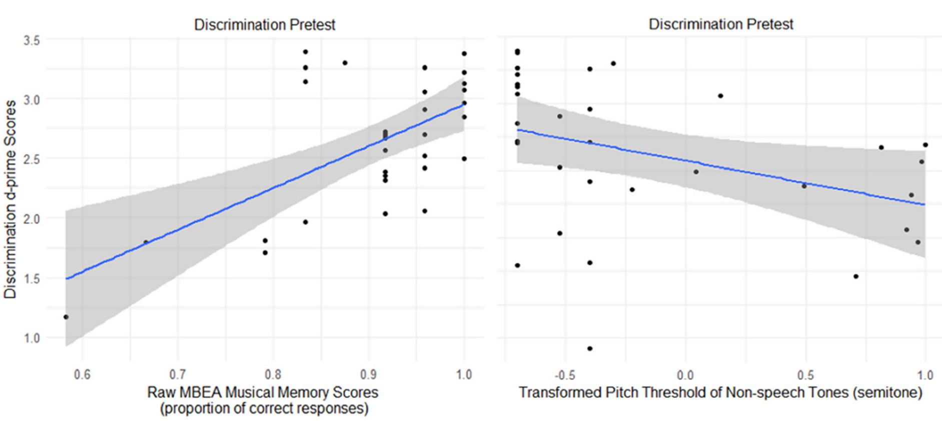

FIG. 1. (Color online) Raw MBEA musical memory scores (left) and transformed nonspeech pitch threshold (right) plotted against the listeners' *d*-prime scores of the discrimination task in the pretest session.

https://doi.org/10.1121/10.0003330

TABLE II. The regression model on *d*-prime scores of the AX discrimination posttest. *B*, standardized coefficients of beta; SE, coefficients standard error.

Variables	Predictors	β	SE	t	p
d-prime scores of AX discrimination posttest	Intercept		0.07	35.56	<0.001
	d-prime scores of pretest	0.81	0.05	7.44	<0.001

including the (*z*-normalized and centered) *d*-prime scores of the AX discrimination pretest as a model covariate/predictor. The final model of the AX discrimination posttest significantly accounted for variance in the *d*-prime scores of the task ($F_{1,30} = 55.35$, $p < 0.001$; adjusted $R^2 = 0.64$). A *post hoc* analysis showed that the power of the model achieved was 0.96. The variance inflation factors of the predictors remaining in the model had values between 1 and 1.5, which indicated a low multicollinearity.

Table II summarizes the regression model on *d*-prime scores of the AX discrimination posttest with pretest *d*-prime scores remaining as the significant predictor. In the model, the effect of the pretest *d*-prime scores independently accounted for a significant portion of the variance after adjusting for the speech pitch threshold ($\beta = -0.23$, $t = -2.02$, $p = 0.05$), nonspeech pitch threshold ($\beta = 0.10$, $t = 0.88$, $p = 0.39$), MBEA-pitch scores ($\beta = 0.13$, $t = 1.01$, $p = 0.28$), MBEA-duration scores ($\beta = 0.01$, $t = 0.04$, $p = 0.97$), and musical memory scores ($\beta = 0.09$, $t = 0.66$, $p = 0.52$). In addition to the pretest *d*-prime scores remaining in the model, the speech pitch threshold is another potential predictor given the marginally significant relationship ($p = 0.05$) between the *d*-prime scores of the AX discrimination posttest and the pitch threshold of speech tones.

Figure 2 shows how the listeners' *d*-prime scores of the AX discrimination posttest are predicted by their pretest *d*-prime scores and speech pitch threshold. Specifically, a higher posttest *d*-prime score after training was associated with a higher pretest *d*-prime score and a lower (more sensitive) speech pitch threshold. There was no significant relationship between the posttest *d*-prime scores and other predictors. The results suggested that the Mandarin listeners' pretest *d*-prime scores, together with their pitch thresholds of speech tones (but not that of nonspeech tones), predicted their discrimination of Cantonese level-tones after the perceptual learning. It is also noted that whereas the pitch threshold of nonspeech tones was predictive in the AX discrimination pretest, the pitch threshold of speech tones was potentially predictive in the AX discrimination posttest.

C. Tone learning: Tone ID

Unlike the AX discrimination task, we do not have a measure of ID performance in the pretest session. The tone-category pairing would have been random before the participants learned to categorize tones in the training session (Chang *et al.*, 2017). To test whether the listeners showed a learning effect in the ID task, two analyses were conducted. The first analysis concerns whether the participants had improved in identifying the novel tone categories at the end of the training session compared to their performance at the beginning of the training session. The second analysis seeks to verify whether the participants performed above chance (i.e., accuracy above 0.33) after the training thereby demonstrating the learning of the tonal categories as intended (Earle and Arthur, 2017; Earle and Myers, 2015).

For the first analysis, a paired-samples *t*-test was conducted on the participants' proportion of correct responses in their first two training blocks and their final two training blocks out of the five blocks in total. The participants' accuracy in the final two blocks (mean = 0.74; SD = 0.13) was statistically higher than their performance in the first two blocks [mean = 0.70; SD = 0.14; $t(31) = -2.45$, $p = 0.02$], indicating that the perceptual training was effective. Whereas the first analysis revealed improvement over the course of the perceptual training,[2] it is unclear whether the participants have mastered the ability to accurately categorize the level-tones as intended. The second analysis serves to verify whether the participants' accuracy was significantly higher than chance. For this purpose, one-sample *t*-tests were conducted on the participants' proportion of correct responses in their ID posttest. The participants' accuracy (mean = 0.64; SD = 0.12) was statistically above chance [$t(31) = 8.80$, $p < 0.001$]. Crucially, their accuracy

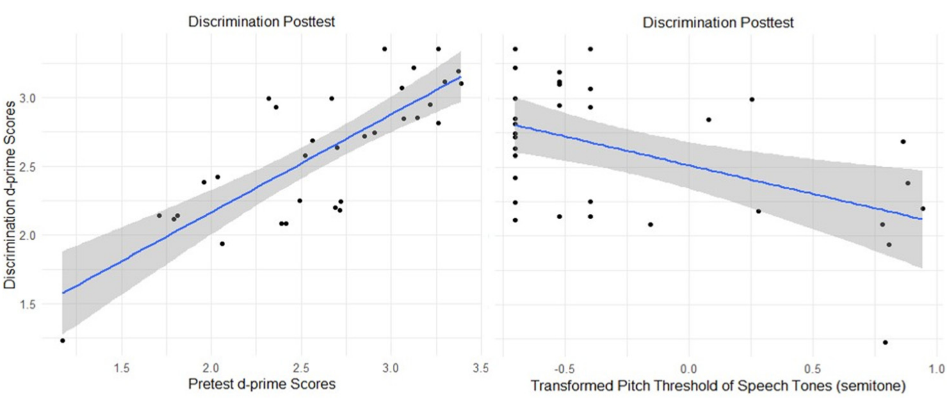

FIG. 2. (Color online) Raw pretest *d*-prime scores (left) and transformed speech pitch threshold (right) plotted against the listeners' *d*-prime scores of the discrimination task in the posttest session.

https://doi.org/10.1121/10.0003330

TABLE III. The regression model on proportion of correct responses in the ID posttest. β, standardized coefficients of beta; SE, coefficients standard error.

Variables	Predictors	β	SE	t	p
Proportion of correct responses of the ID posttest	Intercept		0.02	39.95	<0.001
	d-prime scores of pretest	0.45	0.02	3.22	0.003
	Pitch threshold of nonspeech tones	−0.42	0.03	−3.03	0.005

for the stimuli produced by the trained female speaker (i.e., the stimuli they have heard during training; mean=0.68; SD=0.13; $t(31) = 14.94$, $p < 0.001$) and the new stimuli produced by the untrained female speaker (i.e., the stimuli they did NOT hear during training; mean = 0.60; SD = 0.12; $t\ (31) = 13.11$, $p < 0.001$) were both statistically above chance. The results indicate that the participants learned to categorize stimuli with various pitch heights into the three level-tones accordingly for both the trained talker and the new talker with above-chance accuracy.

The listeners' posttest ID accuracy could also be affected by their pretest discrimination performance to some extent. To determine whether the relationships between the listeners' posttest ID accuracy and the predictors were independent of, or epiphenomenal to, the listeners' pretest discrimination abilities, we, therefore, ran the same regression analysis on the posttest ID accuracy with the same predictors of musical and pitch aptitude. The (z-normalized and centered) *d*-prime scores of the AX discrimination pretest, again, were included as a model covariate/predictor. The final model on the ID posttest significantly accounted for variance in the proportion of correct responses ($F_{2,29} = 15.62$, $p < 0.001$; adjusted $R^2 = 0.49$). A *post hoc* analysis showed that the power of the model achieved was 0.99. The variance inflation factors of the predictors remaining in the model had values between 1 and 1.5, which indicated a low multicollinearity.

Table III summarizes the regression model on the accuracy of the ID posttest with pretest *d*-prime scores and nonspeech pitch threshold as significant predictors. In the model, the effect of pretest *d*-prime scores and nonspeech pitch threshold independently accounted for a significant portion of the variance after adjusting for MBEA-pitch scores ($\beta = 0.26$, $t = 1.98$, $p = 0.057$), MBEA-duration scores ($\beta = 0.15$, $t = 1.14$, $p = 0.27$), musical memory scores ($\beta = 0.16$, $t = 0.95$, $p = 0.35$), and speech pitch threshold ($\beta = -0.20$, $t = -1.13$, $p = 0.27$).

Figure 3 shows how the listeners' posttest ID accuracy was predicted by their *d*-prime scores of the AX discrimination pretest and nonspeech pitch threshold. Specifically, higher ID accuracy after learning was associated with a higher *d*-prime score of the AX discrimination pretest and a lower (more sensitive) nonspeech pitch threshold. While there was a marginally significant relationship between the ID accuracy and MBEA-pitch scores ($p = 0.057$), there was no significant relationship between the ID accuracy and other predictors. The results suggested that the Mandarin listeners' initial discrimination abilities and their sensitivity to nonspeech tones, and possibly also their MBEA-pitch scores to some extent, predicted their ID of Cantonese level-tones after the perceptual learning. It is also noted that while the pitch threshold of speech tones was predictive in the AX discrimination posttest, the pitch threshold of nonspeech tones was predictive in both the AX discrimination pretest and ID test.

To sum up, before the perceptual training, the Mandarin listeners' musical memory abilities and their pitch thresholds of nonspeech tones predicted individual listeners' initial discrimination of Cantonese level-tones. After the perceptual training, the Mandarin listeners' initial discrimination abilities and their pitch thresholds of speech tones predicted their discrimination of Cantonese level-tones. As for the posttest ID accuracy, the participants' initial discrimination abilities and pitch thresholds of nonspeech tones, possibly together with their abilities of processing melodic pitch, accounted for individual variations in the learning outcome.

IV. DISCUSSION

Whereas previous research examined contour-tone learning by learners without experience of a tonal language (Bowles *et al.*, 2016; Dong *et al.*, 2019; Perrachione *et al.*,

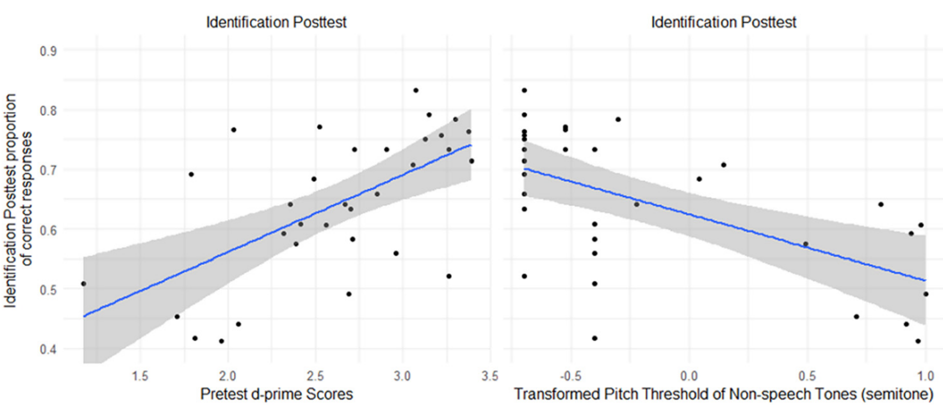

FIG. 3. (Color online) Raw pretest *d*-prime scores (left) and transformed nonspeech pitch threshold (right) plotted against the listeners' proportion of correct responses of the ID in the posttest session.

2011; Sadakata and McQueen, 2014), this study investigated how Mandarin listeners' musical and pitch aptitude modulated their perception and learning of Cantonese level-tone contrasts. In the text below, we first discuss the tone learning effect in the discrimination and ID tasks and then turn to the discussion of the relationship between learners' musical/pitch aptitude and their tone perception/learning performance.

First, the results of posttests suggested that the learners achieved above-chance accuracy of categorizing tones produced by the trained and new talker in the ID tasks. However, the results did not show an effect of learning on the discrimination tasks given that no significant difference was detected between the discrimination pretest and posttest. A plausible explanation is the different nature of training (ID) and assessment (discrimination) tests, which tapped into different aspects of non-native tone perception. The ID task used in the training session (and posttests) tapped into higher levels of phonological encoding of tonal categories (Wang *et al.*, 1999). In contrast, the posttest discrimination tasks, which were designed to test a listener's implicit ability to distinguish tones, might have tapped into relatively low levels of phonetic processing instead. Previous training research found that phonetic processing in sound discrimination did not change much even after multiple ID training sessions (Iverson *et al.*, 2003, 2008, 2012). Moreover, previous studies (Wayland and Li, 2008) suggested a difference between ID training and discrimination training. Specifically, a categorization (or ID) training of tone stimuli might result in an "acquired similarity," that is, a decrease in sensitivity to within-category differences; in contrast, an alternative discrimination training of the same set of stimuli may yield an increase in sensitivity to within-category differences (also see Guenther *et al.*, 1999 for training on non-speech stimuli). In addition, we did not include an ID task in the pretest due to the anticipated poor and potentially random performance (e.g., around chance-level accuracy) of Mandarin listeners in identifying Cantonese level-tones without any training. Nonetheless, future studies may consider including an ID task pretest as a baseline measure to better assess learning-related changes in the ID performance.

More importantly, with regard to the factors predicting the initial perception and learning outcome (Darcy *et al.*, 2015; Golestani and Zatorre, 2009), some musical (e.g., musical memory) and pitch-height processing (e.g., speech or nonspeech pitch threshold) abilities of the Mandarin-speaking learners were found to predict their initial perception and learning of Cantonese level-tones under specific circumstances (pretest or posttest; discrimination or ID). Therefore, the relationship between musical/pitch abilities and the tone perception/learning performance requires further interpretation.

Regarding musical aptitude, we found that incidental musical memory was predictive of the pretest discrimination performance. Different from tone ID, which requires an explicit mapping between an auditory stimulus and a tonal category, tone discrimination is an implicit task using a listener's implicit ability to encode and retain the acoustic cues of the speech signal in short-term memory for phonetic comparison (Earle and Myers, 2014; Wayland and Li, 2008). Although the participants were not instructed to memorize the speech stimuli in the discrimination task, it is not unlikely that their brain would automatically track and retain the speech stimuli that have been presented to them. Likewise, in the incidental musical memory subtest, the listeners were not explicitly instructed to memorize the melodies beforehand but were assessed on their ability to recognize the melodies they have heard and implicitly stored in preceding MBEA subtests (Peretz *et al.*, 2003). There may be potentially shared mechanisms between the retention of auditory musical sequences in the musical memory subtest and lexical tones in the discrimination task. This possibility might be indirectly supported by the effect of implicit statistical learning (i.e., passive exposure to auditory sequences; Caldwell-Harris *et al.*, 2015; Saffran *et al.*, 1996) and indexical learning (i.e., learning non-native speech sounds by playing games cued by auditory stimuli with no explicit rules/instruction; Chan and Leung, 2019; Wiener *et al.*, 2019). To the best of our knowledge, this research is the first to examine the effect of (implicit) musical memory, separated from other dimensions (pitch and duration) of musical aptitude, on non-native tone perception and learning (Chen *et al.*, 2016; Cui and Kuang, 2019). Future research should examine the perception of different types of tones (contour tones vs level-tones) and/or recruit participants with different language backgrounds (tone vs nontonal languages) to further understand the relationship between musical memory and tone discrimination performance.

Moreover, a marginally significant relationship between MBEA-pitch scores (i.e., subtests assessing the learners' sensitivity to changes of pitch scale, contour, or interval in a melody) and the posttest tone ID accuracy was observed. The finding demonstrated a numerical trend with higher abilities of processing musical pitch, predicting better learning outcomes in identifying novel tone categories. Note that the ID posttest included two talkers with different pitch ranges. A previous study has found that Mandarin-speaking participants had difficulty identifying non-native level-tones entangled with talker variability (e.g., a talker's high level-tone could have a similar F0 as another talker's mid/low tone because of between-talker variation in the pitch range; Chang *et al.*, 2017). We speculated that identifying the high, mid, and low level-tones produced by two different speakers in this study requires an ability to estimate an unfamiliar speaker's pitch range. We further speculated that those individuals with better abilities of processing musical pitch presumably outperformed those with lower musical aptitude in calibrating the pitch range of an unfamiliar speaker and, thus, more accurately identified the non-native tones produced by two different speakers in the current study (Leather, 1983). The explanation also echoes the finding of a previous study that English-speaking musicians performed better than nonmusicians in identifying Taiwanese high/low-level-tones produced by multiple speakers with varied pitch ranges (Lee *et al.*, 2014). This numerical trend, while being consistent with previous findings regarding the effect of musical aptitude on tone learning by English-speaking nonmusicians (Li and DeKeyser, 2017), warrants further investigation with a larger sample of Mandarin-speaking L2 learners.

Importantly, the finding on the effect of MBEA musical memory, together with a potential effect of MBEA-pitch scores, demonstrated that (a part of) musical aptitude of Mandarin-speaking learners could predict their initial perception and perceptual learning of novel tones despite their prior experience with native Mandarin tones. Therefore, consistent with findings of previous research on the effect of musical aptitude on *native* tone perception by tonal language speakers (Tang *et al.*, 2016; Wu *et al.*, 2015), this finding tentatively supports a general positive transfer from musical ability to *non-native* tone perception by tonal language speakers (Chen *et al.*, 2016; Cui and Kuang, 2019).

Regarding pitch height processing abilities, we predicted that listeners' pitch thresholds in the speech and nonspeech domains would be both relevant to tone perception and learning. It turned out that the pitch threshold of speech or nonspeech tones differentially predicted the participants' performance in different tests. In the discrimination tasks, the pitch threshold of processing nonspeech discrete tones was predictive of the pretest discrimination. This finding is indicative of Mandarin listeners' sensitivity to subtle pitch height differences at a domain-general auditory level when perceiving non-native (novel) tones before training. In contrast, the threshold of processing discrete speech tones predicted the learners' posttest discrimination of tones. This finding is indicative of their increased sensitivity to higher-level pitch structures, such as contrasts of tonal categories, after training. Overall, the relationship between the listeners' sensitivity to fine-grained pitch height differences and their discrimination of level-tones is consistent with (and complements) the previous finding on contour-tone learning. That is, listeners' pretest sensitivity to pitch contour/height patterns predicts their learning outcome of contour-tone/level-tone words (Bowles *et al.*, 2016; Chandrasekaran *et al.*, 2010; Ingvalson *et al.*, 2013; Perrachione *et al.*, 2011).

In the ID task, unexpectedly, the nonspeech pitch threshold rather than the speech pitch threshold predicted the posttest ID performance, which demands an explanation. This result is generally consistent with the recent finding that listeners' auditory processing of pitch accounted for their individual differences of learning non-native vowels (Kachlicka *et al.*, 2019). As the authors argued, one plausible source of difficulties with learning novel vowels may originate from learners' inability or lower ability in processing pitch at a psychoacoustic level. Following this argument, one possible explanation for the current results is that Mandarin listeners' pitch thresholds of nonspeech tones, compared to their threshold of speech tones, reflected more their auditory processing of pitch. The pitch threshold of nonspeech tones was, thus, a better indicator of learners' abilities in processing pitch at a psychoacoustic level. This finding suggested that a psychoacoustic assessment of pitch height processing abilities prior to perceptual training of novel tones may predict how these individual listeners learned novel tonal contrasts from an unfamiliar language.

Theoretically, the present research makes a good complement to the current L2 tone research, which mostly focuses on the contour-tone learning by nontonal language speakers. Furthermore, the study has generated new knowledge that listeners' individual musical aptitude (musical memory and, to a lesser extent, melodic pitch) plays an important role in accounting for individual differences in the perception of non-native tones even when the participants under study speak a tonal language. For future research, it would be interesting to compare tonal language speakers with nontonal language speakers to investigate whether the effect of musical aptitude on non-native tone perception is larger in one group than it is in the other group (Cui and Kuang, 2019).

The findings of the present study also have pedagogical implications for the learning of lexical tones in the real-world context. Recent decades have seen an influx of Mandarin-speaking immigrants to Hong Kong and it is no trivial matter for them to make the language transition. Although Mandarin is one of the three official languages in Hong Kong, Cantonese, which is used in a wide range of contexts (e.g., TV broadcast and everyday communication), is admittedly the most prominent language in Hong Kong (Matthews and Yip, 2011). Our research with Cantonese tones as the target structure of L2 learning will advance the development of teaching or learning Cantonese as an L2 in Hong Kong and other Pearl River delta areas (Lee, 2020). Specifically speaking, assuming that the goal of Cantonese-tone learning is to build a robust representation of non-native lexical tones, the L2 learners of Cantonese would need to develop the ability to distinguish and categorize the target tones. Therefore, practicing both tone discrimination and ID skills in the perceptual learning paradigm is required for L2 learners to achieve the tone learning outcome (Li and DeKeyser, 2017; Wayland and Li, 2008). Moreover, given the effect of musical aptitude and pitch processing abilities on the L2 tone perception, musical melodies and nonspeech tones can be used as training materials to benefit the learning of L2 tones (e.g., Ireland *et al.*, 2018).

To conclude, complementing existing studies targeting the contour-tone learning from learners without tonal language experience, this study examined the perception and learning of Cantonese level-tones by Mandarin-speaking learners. Our finding suggests an advantage of learners' higher musical aptitude and better pitch (height) processing abilities in L2 tone perception, despite their prior experience with native Mandarin tones. This finding tentatively supports a positive transfer of musical aptitude and pitch sensitivity to the perception of novel level-tone contrasts. Meanwhile, this research also raised questions regarding the underlying mechanism of processing speech vs nonspeech pitch and the potential influence of L2 learners' language background (e.g., tonal and nontonal language speakers) for further research.

ACKNOWLEDGMENTS

This research was partially supported by a *Language Learning* Early Career Research Grant and the Postdoctoral Fellowships Scheme at the Department of Chinese and Bilingual Studies of the Hong Kong Polytechnic University, awarded to Z.Q.; by the Departmental Reward Scheme for Research Publications in Indexed Journals, awarded to C.Z.; and by Grant No. HKRGC-GRF 15606119, awarded to W.S.-y.W.

[1]See http://www.paradigmexperiments.com/ (Last viewed 10 January 2020).
[2]As suggested by an anonymous reviewer, we also examined whether there is improvement in performance over time within each of the ID and discrimination tasks in the pretest and posttest even though no feedback was provided in these tasks. The results showed no difference in performance between the first half and second half of the trials for each task.

Alexander, J. A., Wang, P. C. M., and Bradlow, A. R. (**2005**). "Lexical tone perception in musicians and non-musicians," in *9th Eur. Conf. Speech Commun. Technol.*, pp. 397–400.
Bent, T., Bradlow, A. R., and Wright, B. A. (**2006**). "The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds," J. Exp. Psychol. Hum. Percept. Perform. **32**, 97–103.
Boersma, P., and Weenink, D. (**2018**). "Praat: Doing phonetics by computer (version 6.0.43) [computer program]," http://www.praat.org (Last viewed 8 September 2018).
Bowles, A. R., Chang, C. B., and Karuzis, V. P. (**2016**). "Pitch ability as an aptitude for tone learning," Lang. Learn. **66**, 774–808.
Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. (**1997**). "Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production," J. Acoust. Soc. Am. **101**, 2299–2310.
Caldwell-Harris, C. L., Lancaster, A., Ladd, D. R., Dediu, D., and Christiansen, M. H. (**2015**). "Factors influencing sensitivity to lexical tone in an artificial language: Implications for second language learning," Stud. Second Lang. Acquis. **37**, 335–357.
Chan, R. K. W., and Leung, J. H. C. (**2019**). "Why are lexical tones difficult to learn?," Stud. Second Lang. Acquis. **42**, 33–59.
Chandrasekaran, B., Sampath, P. D., and Wong, P. C. M. (**2010**). "Individual variability in cue-weighting and lexical tone learning," J. Acoust. Soc. Am. **128**, 456–465.
Chang, Y. S., Yao, Y., and Huang, B. H. (**2017**). "Effects of linguistic experience on the perception of high-variability non-native tones," J. Acoust. Soc. Am. **141**, EL120–EL126.
Chen, A., Liu, L., and Kager, R. (**2016**). "Cross-domain correlation in pitch perception, the influence of native language," Lang. Cogn. Neurosci. **31**, 751–760.
Cui, A., and Kuang, J. (**2019**). "The effects of musicality and language background on cue integration in pitch perception," J. Acoust. Soc. Am. **146**, 4086–4096.
Darcy, I., Park, H., and Yang, C. L. (**2015**). "Individual differences in L2 acquisition of English phonology: The relation between cognitive abilities and phonological processing," Learn. Individ. Differ. **40**, 63–72.
Díaz, B., Baus, C., Escera, C., Costa, A., and Sebastián-Gallés, N. (**2008**). "Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language," Proc. Natl. Acad. Sci. U.S.A. **105**, 16083–16088.
Dong, H., Clayards, M., Brown, H., and Wonnacott, E. (**2019**). "The effects of high versus low talker variability and individual aptitude on phonetic training of Mandarin lexical tones," PeerJ **7**, e7191.
Earle, F. S., and Arthur, D. T. (**2017**). "Native phonological processing abilities predict post-consolidation nonnative contrast learning in adults," J. Acoust. Soc. Am. **142**, EL525–EL531.
Earle, F. S., Landi, N., and Myers, E. B. (**2017**). "Sleep duration predicts behavioral and neural differences in adult speech sound learning," Neurosci. Lett. **636**, 77–82.
Earle, F. S., and Myers, E. B. (**2014**). "Building phonetic categories: An argument for the role of sleep," Front. Psychol. **5**, 1192.
Earle, F. S., and Myers, E. B. (**2015**). "Sleep and native language interference affect non-native speech sound learning," J. Exp. Psychol. Hum. Percept. Perform. **41**, 1680–1695.
Erdfelder, E., Faul, F., Buchner, A., and Lang, A. G. (**2009**). "Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses," Behav. Res. Methods **41**, 1149–1160.
Escudero, P., Benders, T., and Lipski, S. C. (**2009**). "Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners," J. Phon. **37**, 452–465.
Farnsworth, P. R. (**1969**). "Tests of musical ability and appreciation: An investigation into the measurement, distribution, and development of musical capacity," PsycCRITIQUES **14**, 306.
Flege, J. E., Bohn, O.-S., and Jang, S. (**1997**). "Effects of experience on non-native speakers' production and perception of English vowels," J. Phon. **25**, 437–470.
Flege, J. E., Takagi, N., and Mann, V. (**1996**). "Lexical familiarity and English-language experience affect Japanese adults' perception of /ɹ/ and /l/," J. Acoust. Soc. Am. **99**, 1161–1173.
Francis, A. L., and Ciocca, V. (**2003**). "Stimulus presentation order and the perception of lexical tones in Cantonese," J. Acoust. Soc. Am. **114**, 1611–1621.
Gandour, J. (**1981**). "Perceptual dimensions of tone: Evidence from Cantonese," J. Chin. Linguist. **9**, 20–36.
Gandour, J. T. (**1983**). "Tone perception in Far Eastern languages," J. Phon. **11**, 149–175.
Golestani, N., and Zatorre, R. J. (**2009**). "Individual differences in the acquisition of second language phonology," Brain Lang. **109**, 55–67.
Guenther, F. H., Husain, F. T., Cohen, M. A., and Shinn-Cunningham, B. G. (**1999**). "Effects of categorization and discrimination training on auditory perceptual space," J. Acoust. Soc. Am. **106**, 2900–2912.
Hao, Y. C. (**2012**). "Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers," J. Phon. **40**, 269–279.
Ho, A. T. (**1976**). "Mandarin tones in relation to sentence intonation and grammatical structure," J. Chin. Linguist. **4**, 1–13.
Ho, O.-Y., Shao, J., Ou, J., Law, S.-P., and Zhang, C. (**2018**). "Tone merging patterns in congenital amusia in Hong Kong Cantonese," in *Proc. Sixth Int. Symp. Tonal Asp. Lang., International Speech Communication*, pp. 134–138.
Ingvalson, E. M., Barr, A. M., and Wong, P. C. M. (**2013**). "Poorer phonetic perceivers show greater benefit in phonetic-phonological speech learning," J. Speech, Lang. Hear. Res. **56**, 1045–1050.
Ireland, K., Parker, A., Foster, N., and Penhune, V. (**2018**). "Rhythm and melody tasks for school-aged children with and without musical training: Age-equivalent scores and reliability," Front. Psychol. **9**, 426.
Iverson, P., Ekanayake, D., Hamann, S., Sennema, A., and Evans, B. G. (**2008**). "Category and perceptual interference in second-language phoneme learning: An examination of English /w/-/v/ learning by Sinhala, German, and Dutch speakers," J. Exp. Psychol. Hum. Percept. Perform. **34**, 1305–1316.
Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (**2003**). "A perceptual interference account of acquisition difficulties for non-native phonemes," Cognition **87**, B47–B57.
Iverson, P., Pinet, M., and Evans, B. G. (**2012**). "Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels," Appl. Psycholinguist. **33**, 145–160.
Jackendoff, R., and Lerdahl, F. (**2006**). "The capacity for music: What is it, and what's special about it?," Cognition **100**, 33–72.
Jiao, L., and Xu, Y. (**2019**). "Whispered Mandarin has no production-enhanced cues for tone and intonation," Lingua **218**, 24–37.
Jongman, A., Qin, Z., Zhang, J., and Sereno, J. A. (**2017**). "Just noticeable differences for pitch direction, height, and slope for Mandarin and English listeners," J. Acoust. Soc. Am. **142**, EL163–EL169.

Kachlicka, M., Saito, K., and Tierney, A. (**2019**). "Successful second language learning is tied to robust domain-general auditory processing and stable neural representation of sound," Brain Lang. **192**, 15–24.
Kim, D., Clayards, M., and Goad, H. (**2018**). "A longitudinal study of individual differences in the acquisition of new vowel contrasts," J. Phon. **67**, 1–20.
Ladd, D. R. (**1996**). *Intonational Phonology* (Cambridge University Press, Cambridge, UK).
Larson-Hall, J. (**2015**). *A Guide to Doing Statistics in Second Language Research Using SPSS and R*, 2nd ed. (Routledge, New York).
Leather, J. (**1983**). "Speaker normalization in perception of lexical tone," J. Phon. **11**, 373–382.
Leather, J. (**1987**). "F0 pattern inference in the perceptual acquisition of second language tone," in *Sound Patterns in Second Language Acquisition*, edited by A. James and J. Leather (De Gruyter Mouton, Berlin), pp. 59–80.
Lee, C.-Y. (**2009**). "Identifying isolated, multispeaker Mandarin tones from brief acoustic input: A perceptual and acoustic study," J. Acoust. Soc. Am. **125**, 1125–1137.
Lee, C.-Y., and Hung, T.-H. (**2008**). "Identification of Mandarin tones by English-speaking musicians and nonmusicians," J. Acoust. Soc. Am. **124**, 3235–3248.
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.
Lee, C.-Y., Lee, Y.-F., and Shr, C.-L. (**2011**). "Perception of musical and lexical tones by Taiwanese-speaking musicians," J. Acoust. Soc. Am. **130**, 526–535.
Lee, C.-Y., Lekich, A., and Zhang, Y. (**2014**). "Perception of pitch height in lexical and musical tones by English-speaking musicians and non-musicians," J. Acoust. Soc. Am. **135**, 1607–1615.
Lee, S. (**2020**). "Pedagogical foci of teaching Cantonese as a second language," Int. J. Curric. Dev. Learn. Meas. **1**, 1–10.
Li, C. N., and Thompson, S. A. (**1977**). "The acquisition of tone in Mandarin-speaking children," J. Child Lang. **4**, 185–199.
Li, M., and DeKeyser, R. (**2017**). "Perception practice, production practice, and musical ability in L2 Mandarin tone-word learning," Stud. Second Lang. Acquis. **39**, 593–620.
Lin, W. C. J. (**1985**). "Teaching Mandarin tones to adult English speakers: Analysis of difficulties with suggested remedies," RELC J. **16**, 31–47.
Liu, F., Xu, Y., Patel, A. D., Francart, T., and Jiang, C. (**2012**). "Differential recognition of pitch patterns in discrete and gliding stimuli in congenital amusia: Evidence from Mandarin speakers," Brain Cogn. **79**, 209–215.
Macmillan, N. A., and Creelman, C. D. (**2004**). *Detection Theory: A User's Guide*, 2nd ed. (Psychology, Mahwah, NJ).
Matthews, S., and Yip, V. (**2011**). *Cantonese: A Comprehensive Grammar*, 2nd ed. (Routledge, London).
Mok, P. K. 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.
Patel, A. D. (**2008**). *Music, Language, and the Brain* (Oxford University Press, Oxford, UK).
Pelzl, E. (**2019**). "What makes second language perception of Mandarin tones hard?," Chin. Second Lang. **54**, 51–78.
Peretz, I., Champod, A. S., and Hyde, K. (**2003**). "Varieties of musical disorders: The Montreal battery of evaluation of amusia," Ann. N.Y. Acad. Sci. **999**, 58–75.
Peretz, I., Gosselin, N., Nan, Y., Caron-Caplette, E., Trehub, S. E., and Beland, R. (**2013**). "A novel tool for evaluating children's musical abilities across age and culture," Front. Syst. Neurosci. **7**, 30.
Perrachione, T. K., Lee, J., Ha, L. Y. Y., and Wong, P. C. M. (**2011**). "Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design," J. Acoust. Soc. Am. **130**, 461–472.
Polka, L. (**1991**). "Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions," J. Acoust. Soc. Am. **89**, 2961–2977.
Qin, Z., and Jongman, A. (**2016**). "Does second language experience modulate perception of tones in a third language?," Lang. Speech **59**, 318–338.
Qin, Z., Tremblay, A., and Zhang, J. (**2019**). "Influence of within-category tonal information in the recognition of Mandarin-Chinese words by native and non-native listeners: An eye-tracking study," J. Phon. **73**, 144–157.
Qin, Z., and Zhang, C. (**2019**). "The effect of overnight consolidation in the perceptual learning of non-native tonal contrasts," PLoS One **14**, e0221498.
Sadakata, M., and McQueen, J. M. (**2014**). "Individual aptitude in Mandarin lexical tone perception predicts effectiveness of high-variability training," Front. Psychol. **5**, 1318.
Sadakata, M., Weidema, J. L., and Honing, H. (**2020**). "Parallel pitch processing in speech and melody: A study of the interference of musical melody on lexical pitch perception in speakers of Mandarin," PLoS One **15**, e0229109.
Saffran, J. R., Aslin, R. N., and Newport, E. L. (**1996**). "Statistical learning by 8-month-old infants," Science **274**, 1926–1928.
Schertz, J., Cho, T., Lotto, A., and Warner, N. (**2015**). "Individual differences in phonetic cue use in production and perception of a non-native sound contrast," J. Phon. **52**, 183–204.
Sun, S. (**1997**). "The development of a lexical tone phonology in American adult learners of standard Mandarin Chinese" (University of Hawaii at Manoa, Honolulu, HI).
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.
Tong, X., Choi, W., and Man, Y. Y. (**2018**). "Tone language experience modulates the effect of long-term musical training on musical pitch perception," J. Acoust. Soc. Am. **144**, 690–697.
Tse, J. K. P. (**1978**). "Tone acquisition in Cantonese: A longitudinal case study," J. Child Lang. **5**, 191–204.
Wagner-Gough, J., and Hatch, E. (**1975**). "The importance of input data in second language acquisition studies," Lang. Learn. **25**, 297–308.
Wang, Y., Spence, M. M., Jongman, A., and Sereno, J. A. (**1999**). "Training American listeners to perceive Mandarin tones," J. Acoust. Soc. Am. **106**, 3649–3658.
Wanrooij, K., Escudero, P., and Raijmakers, M. E. J. (**2013**). "What do listeners learn from exposure to a vowel distribution? An analysis of listening strategies in distributional learning," J. Phon. **41**, 307–319.
Wayland, R. P., and Li, B. (**2008**). "Effects of two training procedures in cross-language perception of tones," J. Phon. **36**, 250–267.
Wiener, S., Murphy, T., Goel, A., Christel, M. G., and Holt, L. (**2019**). "Incidental learning of non-speech auditory analogs scaffolds second language learners' perception and production of Mandarin lexical tones," in *Proc. Int. Conf. Phonetic Sci*., pp. 1699–1703.
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.
Wong, P. C. M., 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.
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.
Xu, G., Zhang, L., Shu, H., Wang, X., and Li, P. (**2013**). "Access to lexical meaning in pitch-flattened Chinese sentences: An fMRI study," Neuropsychologia **51**, 550–556.
Zhang, C. (**2019**). "Tone and music processing in Chinese," in *The Routledge Handbook of Chinese Applied Linguistics* (Routledge, New York), pp. 673–688.
Zhang, C., Shao, J., and Chen, S. (**2018**). "Impaired perceptual normalization of lexical tones in Cantonese-speaking congenital amusics," J. Acoust. Soc. Am. **144**, 634–647.
Zhao, T. C., and Kuhl, P. K. (**2015**). "Effect of musical experience on learning lexical tone categories," J. Acoust. Soc. Am. **137**, 1452–1463.