Effects of musical and linguistic experience on categorization of lexical and melodic tones

Daniel Chang, Nancy Hedberg, and Yue Wang[a)]
Department of Linguistics, Simon Fraser University, Burnaby, British Columbia, V5A 1S6, Canada

(Received 13 June 2015; revised 10 April 2016; accepted 12 April 2016; published online 6 May 2016)

This study investigated the categorization of Mandarin lexical tones and music melodic tones by listeners differing in linguistic and musical experience (English musicians, English non-musicians, and Mandarin non-musicians). Linguistic tonal continua were created from the Mandarin rising to level, and falling to level tones. Melodic continua were created by varying the note D under the context of C and E. The tasks involved tone discrimination and identification. Results revealed that musical training facilitated Mandarin tone categorization, with English musicians' tone identification approximating native Mandarin patterns, being more categorical than English non-musicians'. However, English musicians showed higher discrimination accuracy than Mandarin listeners but not English non-musicians. This suggests that musical experience was not advantageous in discriminating linguistic tonal variations, which requires listeners to ignore subtle physical differences in order to make categorical judgments. Similarly, Mandarin tone experience affected melodic tone identification, with Mandarin non-musicians approximating English musicians, showing more categorical patterns than English non-musicians. In contrast, Mandarin non-musicians' melodic discrimination was the poorest among the three groups, indicating that their experience with linguistic tone categorization may have decreased their sensitivity to fine-grained pitch variations. These results demonstrate bi-directional transfer of pitch proficiency between speech and music as a function of experience.
 [http://dx.doi.org/10.1121/1.4947497]

[TCB] Pages: 2432–2447

I. INTRODUCTION

The interaction between speech and music has been studied extensively, given that the physical realizations of linguistic and musical information share similar acoustic properties, such as fundamental frequency ($F0$, perceptually as pitch) and duration (Besson *et al.*, 2011; Patel, 2011). For instance, tone languages involve the use of $F0$ height and contour shape for differentiating tonal categories (Francis *et al.*, 2008). Likewise, such $F0$ information is used to distinguish different levels of musical tones (Koelsch *et al.*, 1999). Additionally, for both linguistic and musical tone perception, listeners need to form meaningful, discrete categories from an auditory stream (Patel, 2011). If the processing of linguistic and musical information shares such similar physical properties and categorization processes, then it is logical to purport that experience in one domain may facilitate the perception of the other.

A. Effects of musical experience on linguistic tone perception

Empirical studies have revealed that musical ability positively correlates with the ability to discriminate linguistic tonal contrasts in a non-native, unfamiliar tone language (Cooper and Wang, 2012; Delogu *et al.*, 2006; Marie *et al.*, 2011). For instance, in a study examining the discrimination of Mandarin tones by Italian listeners whose native language ($L1$) is non-tonal, Delogu *et al.* (2006) showed that listeners who scored higher on melodic discrimination also achieved greater accuracy in discriminating different Mandarin tones. Similarly, musical aptitude has been shown to be positively correlated with tone-word identification scores by non-tone language listeners (Cooper and Wang, 2012), consistently indicating that musical skills can be transferrable to linguistic tone skills.

Further research demonstrates that musical experience gained from long-term musical training may enhance non-tone-language listeners' sensitivity to specific pitch features in speech. In particular, musical training experience has been shown to facilitate the perception of $F0$ contour (the primary linguistic tonal cue) rather than $F0$ height (assumed to involve calibrating acoustic input based on internally stored $F0$ templates) (Francis *et al.*, 2008; Huang and Johnson, 2011; Lee *et al.*, 2014). For example, while both American musicians and non-musicians perceive tone height in Taiwanese equally well, musicians are more sensitive to tone contour distinctions than non-musicians (Lee *et al.*, 2014). In addition to the perception of acoustic phonetic tonal information, research has examined whether musical training experience facilitates the acquisition of tone words (i.e., identifying the meaning of different lexical items minimally distinguished by phonemic tones). For example, native English musicians compared to non-musicians

[a)]Electronic mail: yuew@sfu.ca

showed greater improvements in Mandarin tone word identification after both groups were trained to perceive Mandarin tone words (Wong and Perrachione, 2007). Moreover, musical experience interacts with linguistic experience to affect tone learning at both phonetic and lexical levels. Cooper and Wang (2012) examined the perceptual learning of Cantonese tone words by recruiting listeners differing in linguistic and musical training experience: musicians with either a tonal (Thai) or a non-tonal (English) $L1$ as well as non-musicians from these $L1$ backgrounds. The results showed that while English musicians were more accurate than Thai non-musicians in distinguishing Cantonese tones based on phonemic tonal information, both groups were more accurate than English non-musicians in identifying the meanings of Cantonese tone words. These findings indicate that long-term exposure to music as well as native tone languages benefits non-native tone word learning. However, in distinguishing non-native tones based on phonetic information, listeners with musical training experience exhibited more enhanced sensitivity to tonal distinctions compared to $L1$ tonal experience.

Electro-physiological research with time-sensitive event-related brain potential (ERP) data has also demonstrated more efficient neural encoding of linguistic tones by musicians relative to non-musicians. Musicians, compared to non-musicians, exhibit larger and earlier ERP responses when discriminating linguistic tonal variations, indicating musicians' enhanced sensitivity to acoustic differences (Besson *et al.*, 2011; Chandrasekaran *et al.*, 2009; Marie *et al.*, 2011). Moreover, brain research coupled with behavioral testing reveals different levels of music-to-language effects in early and later stages of processing. For instance, in Chandrasekaran *et al.* (2009), ERPs indexed by mismatch negativity (MMN) were acquired in association with the discrimination of Mandarin tonal distinctions by native English musicians and non-musicians as well as native Mandarin non-musicians. The ERP results revealed positive effects of experience with both speech and music on early, pre-attentive stage pitch processing, with both Mandarin natives and English musicians showing larger MMN responses than English non-musicians. However, at the later attentive stage, the behavioral results showed that the Mandarin non-musicians were less accurate than both English groups in discriminating $F0$ trajectories that fall within the range of a Mandarin tonal category, suggesting that $L1$ experience with classifying $F0$ variations into categories may have led to decreased sensitivity to subtle within-category $F0$ differences. These findings imply that early stages of pitch processing are sensitive to domain-general experience (e.g., sensory-acoustic experience from speech or music), whereas later stages of processing are driven by domain-specific experience (e.g., $L1$ experience with categorical representations).

In sum, the "music-to-speech" studies demonstrate that musical experience may lead to enhanced behavioral and neural sensitivity to linguistic pitch information, which may further facilitate linguistic tone identification and tone word learning in an unfamiliar tone language. These patterns consistently indicate positive transfer of pitch processing from music to linguistic domains, suggesting that domain-specific pitch exposure may shape fundamental sensory circuitry in a domain-general manner.

B. Effects of linguistic experience on musical tone perception

While musical training has been shown to facilitate tone language learning, there has also been evidence that tone language experience may affect musical perception. Given that linguistic and musical tonal perception both involve modulations of $F0$, it is conceivable that tone language experience may provide an advantage in pitch perception in music. Indeed, studies have reported a facilitative effect of tone language experience on music performance (e.g., Bidelman *et al.*, 2013; Deutsch *et al.*, 2009; Deutsch *et al.*, 2006; Pfordresher and Brown, 2009; Stevens *et al.*, 2011). For example, tone (e.g., Cantonese, Mandarin, Thai) language listeners were shown to outperform non-tone language (e.g., English) listeners in discriminating musical pitch variations (Bidelman *et al.*, 2013; Pfordresher and Brown, 2009) and in detecting pitch contour changes in melodies (Bidelman *et al.*, 2013; Stevens *et al.*, 2011), suggesting that tone language speakers' experience with pitch variations in tones enhances their ability to discriminate musical pitch variations. Moreover, tone language (Cantonese) speakers compared to non-tone language (English) speakers have enhanced tonal memory capacity in music, demonstrating tone language users' enhanced cognitive abilities for musical pitch (Bidelman *et al.*, 2013). Similarly, tone language musicians have been found to perform better in absolute pitch perception than non-tone language musicians, indicating that pitch categorization ability in speech may be carried over to a music context (Deutsch *et al.*, 2006, 2009).

Research has revealed that use of a language-specific perceptual strategy from native tone language experience can also restrict or inhibit music-related pitch processing (Alexander *et al.*, 2011; Peretz *et al.*, 2011). For instance, in Peretz *et al.* (2011), tone (mostly Mandarin) and non-tone (mostly French) language listeners were asked to discriminate five-tone sequences (with each tone synthesized with a piano timbre), with the fourth tone being replaced by a subtle upward pitch or a downward pitch. Results showed that the tone language group performed more poorly than the non-tone language group in downward pitch detection, although the two groups did not differ in upward pitch detection. The failure to detect the subtle downward pitch changes was attributed to the Mandarin listeners' long-term experience with pitch patterns in Mandarin tones, where the falling tone typically involves a larger range $F0$ change compared to the rising tone (Xu, 1994), resulting in their decreased sensitivity to minute downward pitch changes. Likewise, tonal language users' superiority over non-tonal users' in perceiving small pitch excursion differences was absent when detecting pitch excursions smaller than those occurring in their L1 tones (Bidelman *et al.*, 2013). These results imply that if $L1$ tone features are incompatible with the pitch information to be processed in music (e.g., larger range $F0$ change in $L1$ tone than needed for detecting subtle $F0$ change in music), such $L1$ tone influence may be negative. In sum, while long-term exposure to a tone language may facilitate pitch discrimination and categorization in the musical domain, certain $L1$-specific tone experiences may also interfere with musical perception.

C. Bi-directional influence and theoretical accounts

Previous empirical findings demonstrate cross-domain pitch processing for speech and music, pointing to bi-directional influences of pitch-related proficiency between speech and music as a function of experience (Asaridou and McQueen, 2013; Bidelman *et al.*, 2013; Patel, 2011). Such bi-directional influences appear to be associated with enhancement of both lower-level acoustic sensitivity (e.g., $F0$ discrimination) and higher-level cognitive resources (e.g., tonal memory, tonal categorization) in speech and music (Bidelman *et al.*, 2013).

These patterns are in line with the general theoretical framework of pitch processing (e.g., Zatorre and Gandour, 2008), positing an integrated network which incorporates both cue-dependent processes involving acoustic cue encoding of pitch and functional processes involving abstract, category-level representation of pitch. These processes are influenced by both domain-general experience (e.g., experience with acoustic cues or categorization ability commonly occurred in both speech and music) and domain-specific experience (e.g., language-specific or music-specific categorical functions). Specifically, it has been proposed that shared speech and music processes integrate domain-general and domain-specific pitch information at different levels (Patel, 2011; Zatorre and Gandour, 2008). As such, the auditory system initially encodes incoming pitch information, either linguistic or musical, in a domain-general manner, involving processing of sensory-acoustic information. Such information will then be decoded based on prior experience with functional use of pitch, leading to the development of domain-specific pitch categories in specific cognitive domains, such as speech or music. Based on the "shared sound category learning" hypothesis positing common category learning mechanisms across cognitive domains (Patel, 2008), further accounts argue that experience with such domain-specific pitch categories may in turn sharpen domain-general category learning mechanisms involving extraction of abstract rules and regularities, and subsequently facilitate cross-domain category learning (Asaridou and McQueen, 2013; Cooper and Wang, 2012; Kraus and Chandrasekaran, 2010; Patel, 2008, 2011). Moreover, the degree and nature of cross-domain influence depend on specific experience within each domain (Bidelman *et al.*, 2013). According to the OPERA (overlap, precision, emotion, repetition, and attention) model (Patel, 2011, 2014), musical training facilitates speech perception to a greater degree than does speech experience for music perception, because musical training, which involves constantly regulating and monitoring pitch, demands higher precision pitch encoding than required for speech. In terms of the nature of influence, cross-domain transfer effects may either be facilitative or inhibitory depending on if specific pitch experience gained in one domain is compatible with pitch information to be processed in another domain (Asaridou and McQueen, 2013; Bidelman *et al.*, 2013; Peretz *et al.*, 2011).

It should be noted that these theoretical accounts are mainly grounded by the above-reviewed empirical findings from uni-directional research, either "music-to-speech" or "speech-to-music" (Asaridou and McQueen, 2013; Bidelman *et al.*, 2013; Patel, 2011). Although such findings jointly support the bi-directionality accounts, they may not be directly comparable due to differences in musical and linguistic tasks, target stimuli and languages, and/or participant population used in these studies (Asaridou and McQueen, 2013; Bidelman *et al.*, 2013). One recent ERP study (Hutka *et al.*, 2015) is among the very few that examined bi-directional speech and musical influences in a single design. In this research, native English musicians and non-musicians as well as (tonal) Cantonese non-musicians were asked to discriminate pitch ($F0$) differences of musical tones and timbre ($F1$, first formant frequency) contrasts of speech vowels. The results revealed that Cantonese non-musicians were more sensitive to musical pitch information than English non-musicians, and that musicians were superior to non-musicians in speech timbre discrimination, demonstrating bi-directional transfer between speech and music. However, since the study focused on early-stage cortical processing using a discrimination task (which triggers lower-level sensory-acoustic processing), it did not test the characteristics of speech and music transfer at the categorical level. Theoretically, although the shared sound category learning mechanism hypothesis (Patel, 2008) posits shared category learning strategies underlying domain-general sound learning mechanisms, it has not been addressed from the language perspective (Asaridou and McQueen, 2013). Moreover, the different acoustic attributes used for music ($F0$) and speech ($F1$) did not allow direct comparisons of domain-general and domain-specific transfer between speech and music, or the relative weighting of effects from one domain to another as proposed by OPERA (Patel, 2011). Further research is needed to address the theoretical accounts in terms of the integration of speech and music processing at different levels of processing.

D. Categorical perception of tones

One commonly adopted method to examine the domain-general and domain-specific perception of speech is categorical perception because it involves discrimination and identification of sounds both within and across phonemic categories. Categorical perception research on lexical tone has generally shown that native tone language listeners perceive their $L1$ tones in a more categorical manner than non-native tone language listeners (Hallé *et al.*, 2004; Xu *et al.*, 2006; Zheng *et al.*, 2012). Specifically, tone $L1$ listeners have sharper, more distinct category boundaries than non-tonal $L1$ listeners when performing tone identification tasks; on the other hand, non-tonal $L1$ listeners are better than tonal $L1$ listeners at discriminating fine-grained within-category $F0$ differences (Hallé *et al.*, 2004; Liu, 2013; Peng *et al.*, 2010). Furthermore, tonal $L1$ listeners' identification functions exhibit sharper category boundaries for speech tones than for non-speech tones, whereas non-tonal $L1$

listeners' category boundaries are sharper for non-speech tones than for speech tones (Peng *et al.*, 2010; Xu *et al.*, 2006).

The perception of tone categories also reflects differences in the perceptual weighting of tonal features (Huang and Johnson, 2011; Lee *et al.*, 2014). Native listeners with contour tones in their *L*1's (e.g., Mandarin, Thai) attach more importance to *F*0 direction (the primary linguistic tonal cue) than to *F*0 height (assumed to involve calibrating acoustic input based on internally stored *F*0 templates). In contrast, non-tonal *L*1 listeners (e.g., English, Japanese) rely more on *F*0 height than on *F*0 contour.

*L*1 tone background also influences categorical perception patterns. For example, Peng *et al.* (2010) show that listeners whose *L*1 is tonal (Cantonese) are better at differentiating between-category tonal stimuli in a non-native tone language (Mandarin) than detecting within-category *F*0 differences. This is in contrast to non-tonal *L*1 listeners (German) whose discrimination of Mandarin tones appears to be based on psychophysical differences, showing more precise discrimination of subtle *F*0 differences than categorical tonal distinctions. The patterns exhibited by the Cantonese speakers (perceiving Mandarin tones) suggest that listeners' *L*1 (Cantonese) tone experience with classifying *F*0 into categories at the phonemic level can be generalized to detecting tonal boundaries in the non-native language (i.e., Mandarin). Furthermore, tonal *L*1 listeners' perception of non-native tone categories appears to comply with their *L*1-specific tonal patterns. For example, Cantonese and Mandarin listeners show different between-category boundary locations in the discrimination of level-to-falling tone continuum in Mandarin, with Cantonese listeners' perception being biased to falling tones, as compared to Mandarin listeners' (Peng *et al.*, 2010). This is presumably due to the fact that in the same tone space with falling *F*0, Cantonese has two falling tone categories whereas Mandarin only has one. These results indicate that experience with native tones does not necessarily facilitate non-native tone identification accuracy, as specific phonemic and phonetic features between native and non-native tones may differ (So and Best, 2010). For non-tonal *L*1 listeners, the influence of *L*1 appears to arise from acoustic similarities, in that they categorize tones into their *L*1 prosodic categories based on the perceived acoustic similarities in *F*0 information between tones and their native stress, rhythmic, and intonational systems (So and Best, 2011; Yang and Chan, 2010).

Together, these findings demonstrate that, compared to native listeners, non-native, non-tonal *L*1 listeners' tone perception tends to be less categorical, relying more on physical differences or psychophysical boundaries rather than linguistically-relevant aspects. On the other hand, non-native tonal *L*1 listeners may have enhanced categorization ability, although their *L*1-specific experience may also interfere with their perception of target tonal categories.

E. The present study

The present study investigates the transfer between linguistic (Mandarin) and musical (melodic) tones by three groups of listeners differing in their linguistic and musical experience with pitch: native English musicians (EMs) and native English non-musicians (ENs) with no tonal language background, as well as native Mandarin Chinese non-musicians (MNs) whose *L*1 is tonal. Categorical perception tests were administered to examine the discrimination and identification of Mandarin and musical tonal stimuli, including Mandarin tones (rising, falling, level) and musical tones (three-note arpeggios: CCE, CDE, CEE). Re-synthesized tonal exemplars were derived from these tones by creating step-wise *F*0 changes from one tone category to another for Mandarin (rising to level, falling to level) and musical (CCE to CDE, CDE to CEE) stimulus continua.

As this study explores the bi-directional transfer patterns of linguistic and musical tones in a single experimental design, it will provide new evidence for the extant theoretical accounts on speech-music relationship which have been developed mainly based on uni-directional evidence (Asaridou and McQueen, 2013; Bidelman *et al.*, 2013; Patel, 2011). Furthermore, the use of categorical perception tasks allows cross-domain comparisons not only at the sensory-acoustic level (i.e., discrimination of *F*0 differences between tonal exemplars), but also at the abstract, functional level (i.e., tonal category identification) which has not been addressed yet (Asaridou and McQueen, 2013). This study will thus offer new insights on the extant pitch processing theories (Zatorre and Gandour, 2008) by specifying under what circumstances pitch experience may exert domain-general (i.e., shared by speech and music) versus domain-specific (i.e., speech-specific or music-specific) benefits.

Based on the previous findings of the facilitative effects of musical training experience on linguistic tone perception, we hypothesize that, for Mandarin tone perception, native EMs would outperform ENs. Specifically, the EMs are expected to show better discrimination of *F*0 variations than ENs, presumably due to musicians' enhanced sensitivity to subtle *F*0 differences through musical training experience (Cooper and Wang, 2012; Delugo *et al.*, 2006). Furthermore, the EMs and ENs may show better discrimination of *F*0 variations within tonal categories than MNs, since tonal *L*1 listeners' experience with categorical tonal distinctions has been shown to decrease their sensitivity to fine-grained *F*0 variations within tonal categories (Hallé *et al.*, 2004; Peng *et al.*, 2010). For tone identification, the EMs are expected to exhibit more distinct Mandarin tone category boundaries than the ENs, arguably due to musicians' trained ability to classify pitch into categories based on *F*0 contour direction information (Francis *et al.*, 2008; Lee *et al.*, 2014). On the other hand, the EMs' tone identification may still not be on par with the MNs', as language-specific (Mandarin) tonal features unfamiliar to non-native listeners could inhibit their accurate identification (So and Best, 2010; Peng *et al.*, 2010).

For melodic tone, we would expect the MNs' *L*1 tone experience to affect their melodic processing, although the effects may either be positive or negative. Specifically, given the small (4 Hz) *F*0 differences between the melodic stimuli in the current discrimination task, we predict that the MNs would be inferior to EMs and ENs in melodic tone discrimination, since the MNs' *L*1 experience with categorical tonal distinctions has been shown to decrease their sensitivity to subtle $F0$ differences (cf. Bidelman *et al.*, 2013; Peng *et al.*, 2010). For melodic identification, the MNs would perform better than the ENs due to their $L1$ tone experience with pitch categorization (cf. Bidelman *et al.*, 2013). However, they may still not be on par with the EMs, if their $L1$ tonal features were incompatible with melodic tonal features (cf. Bidelman *et al.*, 2013; Peretz *et al.*, 2011).

Based on the OPERA hypothesis (Patel, 2011), we expect acoustic and categorical level of facilitation from music to speech, but categorical level facilitation from speech to music, since musical training enhances both sensitivity to subtle acoustic variation and categorization ability whereas linguistic experience primarily shapes categorical distinctions. These patterns are also expected to support the pitch processing theories proposed by Zatorre and Gandour (2008), in that bi-directional transfer effects involve an integration of domain-general processes (where common cognitive abilities, e.g., categorization, can be transferrable across linguistic and musical domains) and domain-specific processes (when language-specific and melodic-specific features are not compatible).

II. METHODS

A. Participants

A total of 55 young adult participants (mean age: 26 years) were recruited from the Simon Fraser University graduate and undergraduate student population, including 19 native EMs, 19 native ENs, and 17 native MNs. The native EMs (7 male, 12 female) and ENs (9 male, 10 female) were all born and raised in Canada with no previous knowledge of a tone language. The musicians were defined as those having 5 or more consecutive years (mean: 11 years; range: 5–16 years; standard deviation, $SD = 4.7$) of formal training with a western-style instrument (e.g., piano) and having played the instrument in the past 5 years on a regular basis as of the time of the experiment (following the previously established inclusion criteria for musicians, e.g., Bidelman *et al.*, 2013; Cooper and Wang, 2012; Wong and Perrachione, 2007). Their initial age of formal musical training was 11 years old ($SD = 3.4$). Thirteen of the musicians reported that piano had been the major instrument during their musical training, while the remaining six participants reported having violin, clarinet, guitar, percussion, or flute as the major musical instrument. In contrast, the ENs and MNs had an average of 1 year (maximum: 3 years) of musical training experience, and none of them had any formal musical training within the past 5 years (Cooper and Wang, 2012). The MNs (12 male, 5 female) were all born and raised in China or Taiwan with no other tone language experience or only minimal proficiency in a second tone language (e.g., Cantonese or Taiwanese). They had been residing in Canada for less than 6 years at the time of testing, arriving in the country at 20 years of age on average ($SD = 5$).

B. Stimuli

For Mandarin tone categorical perception testing, three mono-syllabic Mandarin real words differing in tone produced by a 28-year-old male native Mandarin Chinese speaker were selected as the templates of the end-point tonal tokens used to re-synthesize the rising-level and falling-level tone continua, including "*yi1*" (with the high-level tone, Tone 1, meaning "one"), "*yi2*" (with the mid-rising tone, Tone 2, meaning "move"), and "*yi4*" (with the high-falling tone, Tone 4, meaning "meaning"). These target words were recorded in a sound-attenuated recording booth at the Language and Brain Lab, Simon Fraser University, using a Shure KSM109 microphone (Shure Inc.) and a Presonus Firebox audio interface (PreSonus Audio Electronics, Inc.) at a 44.1 kHz sampling rate. The selection of these stimuli was based on multiple repetitions from the speaker in order to select the ones with comparable tone duration (within 10% variance across the three tones) and comparable $F0$ onset and offset (within 5 Hz variance for the onset/offset of the level tone, the onset of the falling tone and offset of the rising tone; and within 5 Hz variance for the onset of the rising tone and offset of the falling tone). These three naturally produced tone words were evaluated by two phonetically trained listeners and rated as good exemplars of Mandarin tones in their respective categories.

Critical $F0$ values of the three selected tone words were adjusted in order to provide the endpoint values for re-synthesizing the stimuli in the tone continua. First, the onset and offset $F0$ of the level tone, the offset $F0$ of the rising tone, and the onset $F0$ of the falling tone were fixed at 152 Hz, which is the averaged value of these points in the original *yi1*, *yi2*, and *yi4* productions. Likewise, the onset $F0$ of the rising tone and the offset of the falling tone was fixed to 128 Hz. The durations of the tones were fixed to 323 ms for the rising-level continua and 294 ms for the falling-level continua based on the corresponding averaged values. Additionally, since the rising and falling tones each involve an initial period of flat $F0$ before they rise or fall, respectively, the turning points where the $F0$ changes from flat to rising or falling (cf. Moore and Jongman, 1997; Peng *et al.*, 2010) were identified based on the original productions of these two tones. For the rising tone, the turning point is at 30% of the entire duration, involving 97 ms of flat $F0$ followed by 226 ms of rising $F0$; for the falling tone, the turning point is at 10%, involving 32 ms of flat $F0$ followed by 262 ms of falling $F0$.

Based on these features of the end-point tone tokens, the tonal continua for the categorical perception tests were created using the previously established PSOLA (Pitch Synchronous Overlap and Add) method (Peng *et al.*, 2010; Xu *et al.*, 2006; Zheng *et al.*, 2012), as displayed in Fig. 1. Two types of tone continua were developed: rising-level and falling-level (modeled after Peng *et al.*, 2010). For each tone pair type, two continua were created, one from each tone as the starting point. For instance, for the rising-level pair, a rising-to-level continuum was created using the rising tone as the starting point with step-wise $F0$ changes toward the level tone, and a level-to-rising continuum was created using the level tone as the starting point. Developing the same type of continua (rising-level) from both directions (rising-to-level, level-to-rising) was to ensure that the tonal features of each tone of the tone pair were captured, thus avoiding

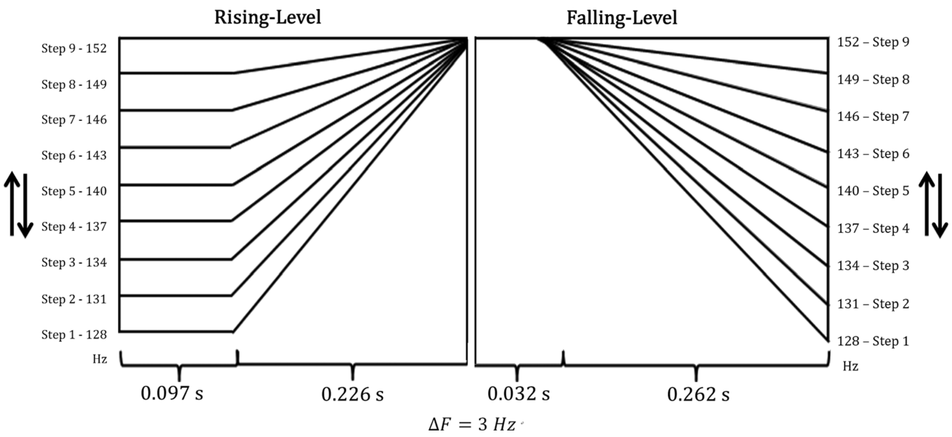

FIG. 1. Mandarin rising-level and falling-level tone continua, created using the PSOLA method based on natural productions of the Mandarin real word syllables *yi1* (Tone 1, level tone), *yi2* (Tone 2, rising tone), and *yi4* (Tone 4, falling tone). Each continuum involves 9 steps with 3 Hz *F*0 increments. The onset and offset *F*0, duration, and turning point values have been normalized based on the original natural productions of these syllables. The two directional arrows by the side of each type of continua are to illustrate that one continuum was created from each direction, as a counterbalancing measure to ensure that the tonal features of each tone of the tone pair were captured.

any potential bias in favor of the tone that a continuum was created from. Likewise, for the falling-level pair, two tone continua were developed, one from each direction. Thus, in total, there were four tone continua: rising-to-level, level-to-rising, falling-to-level, level-to-falling.

The difference between each of the two exemplars in a continuum was 3 Hz (cf. Peng *et al.*, 2010), resulting in nine steps (nine tonal exemplars) including the end-point tones for each continuum. The 3 Hz-interval was used because it has been claimed to be the smallest just noticeable difference (JND) for tonal differences (Harrison, 1996; Liu, 2013), and the selection of this small interval allowed testing of maximal sensitivity to fine-grained pitch changes, especially for the musically trained participants. Last, the intensity of all the stimuli was normalized to 65 dB (Peng *et al.*, 2010; Zheng *et al.*, 2012). In total, 36 tokens (9 steps × 4 continua) were created. The re-synthesized (level, rising, and falling) end-point tone tokens were evaluated by two native Mandarin phoneticians. All were correctly identified and judged as good tone exemplars in Mandarin.

For melodic tone categorical perception testing, a musical tone continuum was created from a three-note arpeggio, generated by Garageband (Bidelman *et al.*, 2011). To control for familiarity, a universally familiar arpeggio was used, involving C4 (261.6 Hz), D4 (293.6 Hz), and E4 (329.6 Hz). The C-D-E continuum was developed with a mis-tuned D4, ranging from 261.6 Hz (equivalent to C4) to 329.6 Hz (equivalent to E4), with each mis-tuned D4 differing in 4 Hz (Harrison, 1996). This results in an 18-step continuum, 9 steps from C4 to D4, and 9 from D4 to E4. Each note (C, D, E) is 100 ms in duration, resulting in a set of 300 ms arpeggios.

Three-note arpeggios were chosen because a melodic context must be provided for the non-musicians to label the melodic tones. Although absolute or relative pitch had been used in previous research to examine the effects of linguistic experience on musical pitch perception, such stimuli were designed for musicians (Deutsch *et al.*, 2006, 2009). The three-note arpeggios used in the present study allow both musicians and non-musicians to discriminate and identify pitch differences in a melodic context. Similar (five-note) melodic contours have been used previously (Galvin *et al.*, 2007). These manipulations of the melodic continuum also reflect an effort in the design to match the melodic and Mandarin tone stimuli in that the three-note arpeggios resemble the pitch contours in Mandarin tones. Particularly, the increasing/decreasing increments of note D in the context of C-D-E create a series of rising melodic contours resembling rising tone contours in Mandarin. Additionally, the melodic and Mandarin tonal exemplars match in terms of duration (both around 300 ms) as well as variation in *F*0 between each of the exemplars in a continuum (3–4 Hz). Figure 2 illustrates the melodic tone continuum.

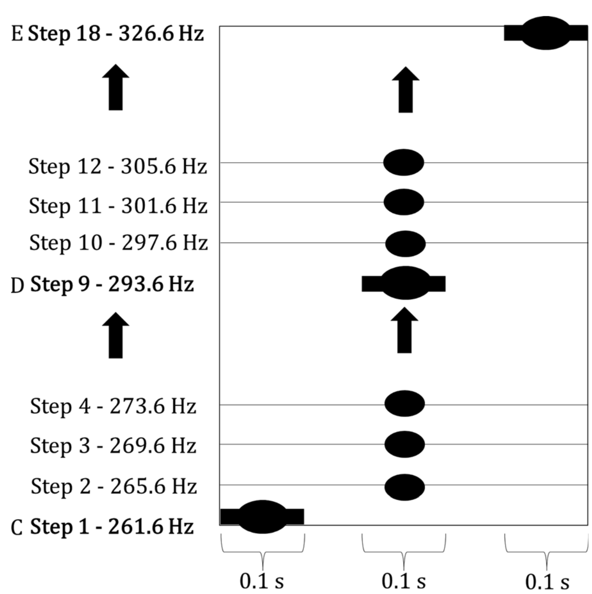

FIG. 2. Melodic tone continuum created from a three-note arpeggio involving C4, D4, and E4. The C-D-E continuum involves a set of mis-tuned D4, each differing in 4 Hz, ranging from C4 to E4. This results in an 18-step continuum, 9 steps from C4 to D4, and 9 from D4 to E4. Each arpeggio is 300 ms in duration (100 ms for each note).

C. Procedures

For categorical perception, first, an AX discrimination task was included for both the linguistic and musical stimuli. For the Mandarin tone condition, there were 25 pairs for each 9-step continuum, including (1) nine "same" pairs (one for each of the nine steps, e.g., step1–step1, step2–step2), and (2) 16 different pairs (with 8 combinations, e.g., step1–step2, step2–step3,…, step8–step9; and the reverse pairs, e.g., step2–step1, step3–step2, etc., as a counterbalancing measure to avoid order effects in an AX pair). For the melodic tone condition, there were 18 same pairs (1 from each of the 18 steps in the continuum, e.g., [CX_1E, CX_1E], with X being the varied D note) and 34 different pairs (with 17 combinations and the reverse pairs, e.g., [CX_1E, CX_2E]), resulting in a total of 52 pairs. For both conditions, each pair was presented three times. In total, there were 300 Mandarin tone pairs (25 pairs $\times$ 3 repetitions $\times$ 4 continua) and 156 melodic tone pairs (52 pairs $\times$ 3 repetitions). The participants were instructed to judge whether the two tones in a pair sounded the same or different. Adjacent pairs (e.g., step1–step2) were used in both the Mandarin tone and melodic tone discrimination tasks since the one-step (3 or 4 Hz) difference has been claimed to be the JND for tone perception (Liu, 2013), thus allowing listener-group comparisons based on their finest sensitivity to tonal differences, especially given that musicians were involved. The inter-stimulus interval (ISI) for the AX discrimination pairs was set at 250 ms. This interval has been claimed to be an appropriate interval for AX speech sound discrimination tasks, since a shorter (than 200 ms) ISI may result in mutual masking of the two stimuli in an AX pair, while a longer (than 300 ms) ISI may exceed the auditory memory life span of the first stimulus in the AX pair (Gerrits and Schouten, 2004; Pisoni, 1973). The Mandarin tone and melodic tone conditions were presented in separate blocks, and for the Mandarin task, the rising-level and falling-level pairs were also presented in separate blocks. The stimuli were presented in a randomized order within each block.

In addition to the discrimination task, a 2-way-alternative-forced-choice (2-AFC) identification task was administered for the linguistic as well as the musical conditions. For the Mandarin tone condition, after a tonal exemplar from one of the continua was displayed, listeners were asked to identify the tone from two response options (e.g., for the rising-level continua, rising tone, or level tone; for the falling-level continua, falling tone, or level tone). There were 9 stimuli per continuum, each repeated 3 times, resulting in a total of 108 stimuli for the 4 tonal continua. The melodic tone identification involved a contextual melody categorization task. For each trial, a three-note arpeggio (CXE) was presented, where X is the varied D note. To maintain a 2-AFC identification task that would match the Mandarin tone identification task, the melodic continuum was separated into two different sets: (1) CCE to CDE, and (2) CDE to CEE. For each trial, participants were asked to determine whether a given arpeggio (CXE) resembles Do-Do-Mi (C-C-E) or Do-Re-Mi (C-D-E) in the first set, or resembles Do-Re-Mi (C-D-E) or Do-Mi-Mi (C-E-E) in the second set. There were 27 stimuli in the "CCE to CDE" condition (9 steps $\times$ 3 repetitions = 27), and 30 stimuli in the "CDE to CEE" condition (10 steps $\times$ 3 repetitions = 30). The two sets of Mandarin tone and two sets of melodic tone stimuli were presented in separate blocks, with the order of stimulus presentation randomized in each block.

All participants completed both the discrimination and identification tasks in the Mandarin tone and melodic tone conditions. All the participants performed the two tasks in the same order, discrimination and then identification, to avoid any potential (tonal category) familiarity effects on the performance of the discrimination task due to prior exposure to category labelling from the identification task, especially for the English listeners. However, the order of presentation of the Mandarin and melodic tone blocks as well as the order of the two sub-blocks of stimuli within each Mandarin or melodic tone condition (rising-level, falling-level; CCE-CDE, CDE-CEE) were counter-balanced across participants. For each discrimination or identification trial, the participants had a maximum of 2 s to give a response.

Prior to each task, the participants received a short 1–2 min familiarization and practice session in order to be familiarized with the stimuli and task. The practice stimuli for the Mandarin tone discrimination task included two same pairs and three "different" pairs randomly selected from the set of discrimination stimuli. Participants completed two repetitions of these stimuli presented in a randomized order, with the manner of presentation of each trial (e.g., ISI) as well as the response task being comparable to those used in the actual discrimination test. The practice stimuli for the Mandarin tone identification task included three repetitions of the three natural tone words (/yi/ with level, rising, and falling tones) used to develop the tone continua. The manner of presentation and task were comparable to those used in the actual identification test. Prior to the practice task, the English participants were also presented with these words with feedback given as to which tone they just heard (level, rising, falling) to ensure that they were able to label the tones. The practice sessions for melodic tone discrimination and identification were administered in a comparable fashion, with the practice stimuli randomly selected from the melodic stimulus set. In total, the experiment lasted a maximum of 1.5 h, including practice, instructions, as well as a short break in between each of the blocks.

D. Analysis

1. Discrimination accuracy

A discrimination accuracy score (P) was computed based on the following Eq. (1), as used in previous studies (e.g., Peng *et al.*, 2010; Xu *et al.*, 2006):

$$P = P(``S''|S) * P(S) + P(``D''|D) * P(D), \qquad (1)$$

where $P(``S''|S)$ refers to the percentage of Same responses of all the same pairs, whereas $P(``D''|D)$ refers to the percentage of Different response among all the different pairs. $P(S)$ and $P(D)$ refer to the probabilities of the Same (AA, BB) and Different (AB, BA) trials in a unit, which are both 50%. This equation controls for response bias. The score ranges from 0 to 1, indicating 0% to 100% correct discrimination, respectively. Based on this equation, the discrimination scores for each pair in each continuum were computed for each participant and subsequently submitted for analyses of variance (ANOVAs) to examine group differences.

2. Identification function

Following previously established models (e.g., Xu *et al.*, 2006), a generalized linear model was used to investigate the sharpness of the category boundary and boundary location in the tone identification tasks. The model involves logistic regression to examine how the binary classification of a sound (e.g., Sound 1 or Sound 2) fits into a sigmoid curve (or a categorical curve) in terms of the category boundary sharpness and location. The identification scores were computed based on the following Eq. (2):

$$\log_e[P_I/(1\text{–}P_I)] = b_0 + b_1 x, \tag{2}$$

where P_I is the probability for the identification options provided to the participants (e.g., Sound 1 or Sound 2), and is set to be 0.5 (50%), because there are two response options. The variable "x" refers to a step number in a continuum, indicating the potential categorical boundary for a continuum. The variable "b_1" refers to the slope of the boundary between the two categories in a continuum, considered the sharpness of the categorical boundary in the continuum. The variable "b_0" refers to the intercept of the regression equation. Therefore, to compute the categorical boundary X_{cb} based on the identification responses gathered from each listener for each group, the following Eq. (3) is derived from Eq. (2):

$$X_{cb} = \text{–}b_0/b_1. \tag{3}$$

The following predictor variables were analyzed using the generalized linear model: Group (EM, EN, MN), Step (Mandarin: Step 1 to Step 9; Music: Step 1 to Step 9; Step 9 to Step 18), and (for Mandarin tone only) Direction (for the rising-level continua: rising-to-level and level-to-rising; for the falling-level continua: falling-to-level and level-to-falling). The slope (category boundary sharpness) and the category boundary position at the 50% crossover point for each continuum and each participant were computed, and then submitted to subsequent ANOVAs to compare group differences.

III. RESULTS

A. Categorical perception of Mandarin tone

1. Mandarin tone discrimination

For each type of tone continuum (rising-level, falling-level)[1], a mixed three-way ANOVA was conducted, with Group (EMs, ENs, MNs) as a between-subject factor, and Tone pair (Pair1–2, Pair2–3, Pair3–4, Pair4–5, Pair5–6, Pair6–7, Pair7–8, Pair8–9, composed of all the adjacent steps from Step 1 to Step 9), Direction of continuum[2] (for the rising-level continua: rising-to-level and level-to-rising; for the falling-level continua: falling-to-level and level-to-falling) as repeated measures. The measurement for the discrimination task was "discrimination accuracy" as described previously, ranging from 0 to 1, with 0 being no correct discrimination, 0.5 being chance level (50%), and 1 being 100% correct discrimination for all the pairs. As the current study focused on group comparisons, only the significant interactions involving group were analyzed further.

For the rising-level continuum, a marginally significant main effect of Group was obtained [F(2, 816) = 2.9, $p = 0.057$], indicating a tendency of group differences in the Mandarin tone discrimination across pairs and directions. *Post hoc* Tukey HSD p-adjusted pair-wise comparisons among the groups revealed that, across tone pairs, EMs (mean value, $M = 0.54$) performed marginally better than MNs ($M = 0.52$, $p = 0.057$). ENs ($M = 0.53$) did not differ significantly from either EMs ($p = 0.880$) or MNs ($p = 0.153$). Furthermore, a significant main effect of Tone pair was found [$F(7, 816) = 0.7$, $p < 0.0001$]. *Post hoc* Tukey HSD p-adjusted pair-wise comparisons showed that the accuracy for Pair5–6 ($M = 0.60$, $p < 0.0001$) was significantly higher than that for the rest of the pairs which did not differ significantly ($M = 0.52$, $p > 0.05$), indicating the discrimination peak location at Pair5–6 (i.e., between Step 5 and Step 6), presumably the location of category boundary. There was no significant main effect of Direction [$F(2, 816) = 0.2$, $p = 0.639$]. No significant interactions were observed for Group × Tone pair [$F(14, 816) = 0.6$, $p = 0.875$], Group × Direction [$F(2, 816) = 0.5$, $p = 0.618$], or Group × Tone pair × Direction [$F(14, 816) = 0.3$, $p = 0.994$]. Furthermore, to determine whether there was any group difference at the discrimination peak location, a one-way ANOVA using Group as the between-subject factor and discrimination score at Pair5–6 (peak) as the dependent variable was conducted. The results showed no significant effect of Group at the discrimination peak [$F(2, 105) = 0.1$, $p = 0.909$]. Figure 3(a) displays the rising-level discrimination scores for the three groups as a function of tone pair (step).

For the falling-level continuum, a significant main effect of Group was obtained [$F(2, 816) = 3.9$, $p = 0.020$]. *Post hoc* Tukey HSD p-adjusted pair-wise comparisons indicated that, across tone pairs, EMs ($M = 0.55$) performed significantly better than MNs ($M = 0.52$, $p = 0.021$). ENs ($M = 0.53$) did not differ significantly from either EMs ($p = 0.094$) or MNs ($p = 0.792$). A significant main effect of Tone pair was also found [$F(7, 816) = 10.0$, $p < 0.0001$], with *post hoc* Tukey HSD p-adjusted analyses showing that Pair5–6 ($M = 0.64$, $p < 0.0001$) was more accurately discriminated than were the rest of the pairs, which did not differ ($M = 0.52$, $p > 0.05$). There was no significant main effect of Direction [$F(1, 816) = 0.9$, $p = 0.353$], or significant interactions for Group × Tone pair [$F(14, 816) = 1.0$, $p = 0.459$], Group × Direction [$F(2, 816) = 0.7$, $p = 0.113$], or Group × Tone pair × Direction [$F(14, 816) = 1.5$, $p = 0.113$]. A further one-way ANOVA using Group as the between-subject factor and discrimination score at Pair5–6 as the dependent variable was conducted, showing no significant group difference at the discrimination peak

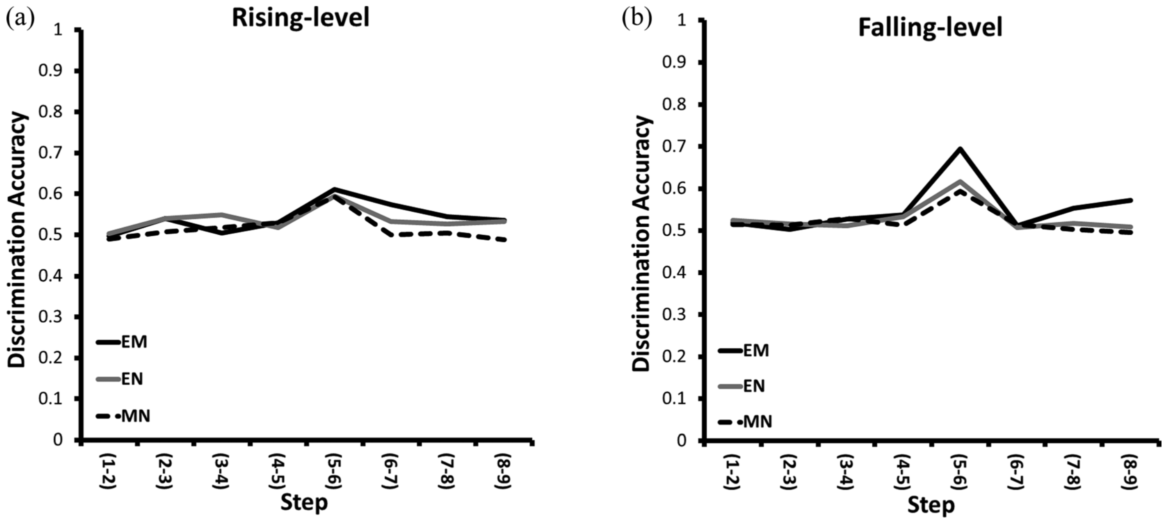

FIG. 3. Mandarin (a) rising-level and (b) falling-level tone discrimination curves for each group (EM, English musicians; EN, English non-musicians; MN, Mandarin non-musicians). The *y* axis shows the discrimination scores ranging from 0 to 1, with 1 being perfect (100% correct) discrimination of the tone pairs and 0.5 being chance level. The *x* axis shows the discrimination pairs (steps) along the rising-level and falling-level continua, respectively.

[$F(2, 105) = 2.5$, $p = 0.087$]. Figure 3(b) displays the falling-level discrimination scores for the three groups as a function of tone pair (step).

In sum, the Mandarin tone discrimination scores showed, across pairs, an overall tendency for the EMs to outperform MNs but not ENs in both rising-level and falling-level continua. However, there was no group difference in the location of discrimination peak (category boundary), all being at Pair5–6 (between Step 5 and Step 6); nor was there any group difference in the discrimination scores at the discrimination peak.

2. Mandarin tone identification

The Mandarin tone logistic identification functions as well as the corresponding discrimination curves are displayed in Fig. 4. The identification curves plot the percent responses of each of the tones as a function of step (1 to 9)

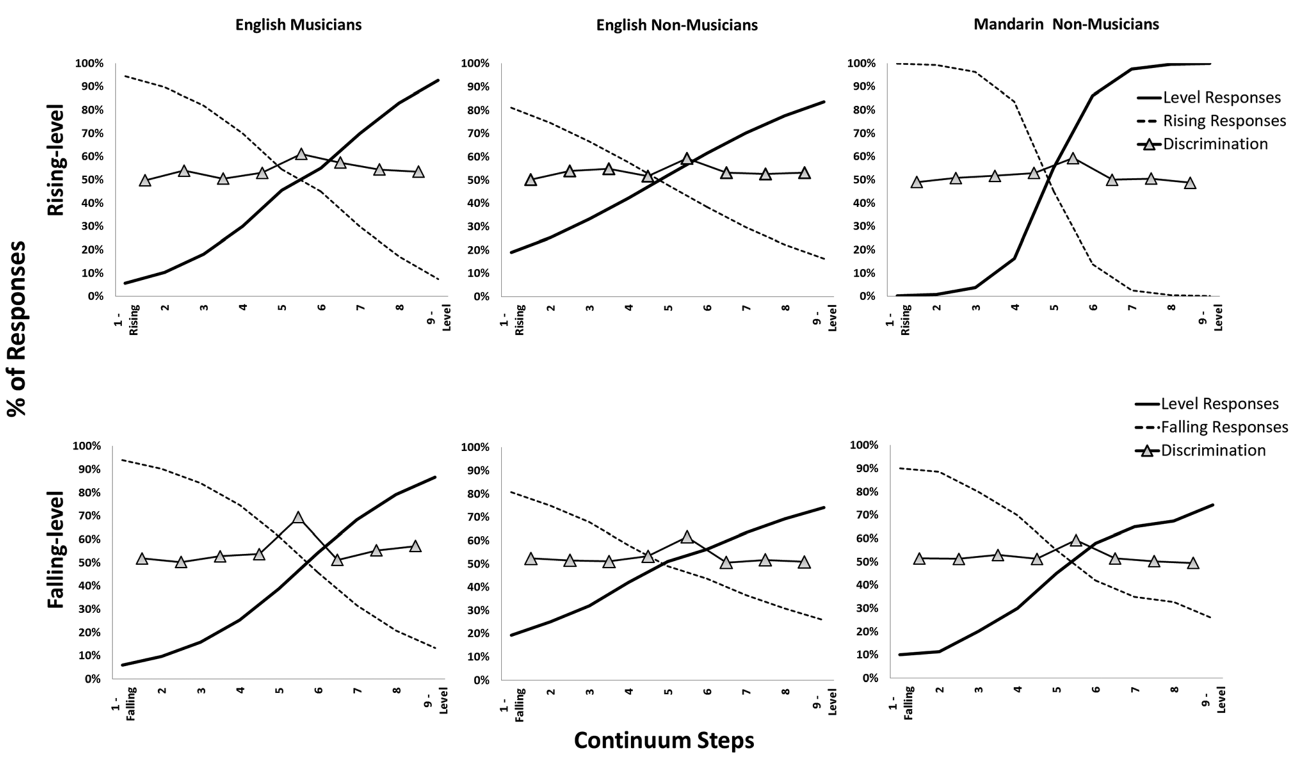

FIG. 4. Mandarin tone logistic identification functions and discrimination curves for each type of continua (rising-level, falling-level) and each group [English musicians (EMs), English non-musicians (ENs), Mandarin non-musicians (MNs)]. The identification curves plot the percent responses of each of the tones as a function of step (1 to 9) along each of the tone continua, with Step 1 being the rising or falling end and Step 9 being the level end.

along each of the tone continua. The identification results in terms of the category boundary sharpness (steepness of slope) and boundary location are derived from the generalized linear model analysis described previously. For slope, a higher value means a greater degree of steepness, and thus a sharper category boundary, indicating that perception is more categorical. The boundary location ranges from 1 to 9, corresponding to the 9 steps in each of the tone continua, in which 1 represents the rising end or the falling tone end, and 9 represents the level tone end. For instance, a boundary location value of 8.6 is closer to the level end, indicating that there are more rising or falling tone responses than level tone responses.

Figure 5 presents the category boundary sharpness (slope steepness) comparisons among the groups for the rising-level and falling-level continua. A two-way ANOVA using Group as the between-subject factor and Direction as the repeated measures was conducted for each type of continua (rising-level, falling-level), revealing a significant main effect of Group for both the rising-level [$F(2, 208) = 10.9$, $p < 0.0001$] and falling-level [$F(2, 208) = 6.8$, $p = 0.002$] continua. For the rising-level continua, *post hoc* Tukey HSD p-adjusted pair-wise comparisons indicated that MNs showed a significantly sharper category boundary than both EMs[3] ($p = 0.014$) and ENs ($p < 0.0001$). Furthermore, EMs showed a significantly sharper boundary than ENs ($p = 0.003$). For the falling-level continua, both MNs ($p = 0.002$) and EMs ($p = 0.015$) showed a significantly sharper boundary than ENs. The sharpness of the category boundary did not differ significantly between MNs and EMs ($p = 0.776$). No significant Group $\times$ Direction interactions were observed for either type of continua [Rising-level: $F(2, 208) = 2.2$, $p = 0.114$; Falling-level: $F(2, 208) = 1.3$, $p = 0.267$].

Figure 6 displays the category boundary location comparisons among the groups for the rising-level and falling-level continua. A two-way ANOVA using Group as the between-subject factor and Direction as the repeated measures was conducted for each type of the continua. For the rising-level continua, there was no significant main effect of Group [$F(2, 95) = 0.8$, $p = 0.463$], showing an averaged category boundary at Step 5 ($M = 5.0$). No significant interaction of Group $\times$ Direction was observed [$F(2, 95) = 0.1$, $p = 0.918$].

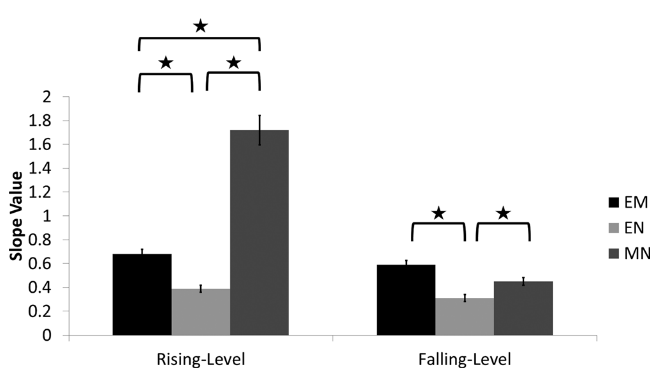

FIG. 5. Mandarin tone slope (category boundary sharpness) comparisons from the logistic identification functions for each type of continua (rising-level, falling-level) and each group (EM, English musicians; EN, English non-musicians; MN, Mandarin non-musicians). Error bars indicate standard error.

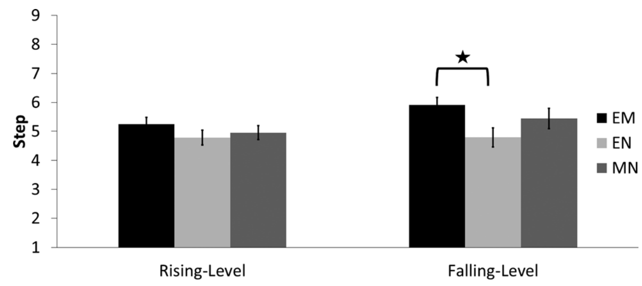

FIG. 6. Mandarin tone category boundary location (step at 50% cross-over) comparisons from the logistic identification functions for each type of continua (rising-level, falling-level) and each group (EM, English musicians; EN, English non-musicians; MN, Mandarin non-musicians). Error bars indicate standard error.

For the falling-level continua, on average the boundary location across groups and direction was between Steps 5 and 6 ($M = 5.4$), in favor of falling tone responses. However, a significant Group effect was obtained [$F(2, 72) = 3.6$, $p = 0.03$]. *Post hoc* Tukey HSD p-adjusted pair-wise comparisons indicated that the EMs' boundary location occurred at a higher step than ENs' ($p = 0.024$), indicating more falling responses for the musicians compared to the non-musicians. There was no location difference between EMs and MNs ($p = 0.541$), or between ENs and MNs ($p = 0.365$). No significant Group $\times$ Direction interaction was obtained [$F(2, 72) = 2.5$, $p = 0.806$].

In sum, the Mandarin tone identification results revealed that for both types of continua (rising-level, falling-level), the EMs followed the native Mandarin patterns showing a sharper category boundary than the ENs, indicating that the EMs' Mandarin tone identification was more categorical than the ENs'. Overall, the category boundary locations were similar across groups, except that for the falling-level condition, the EMs gave slightly more falling-tone responses compared to the ENs.

B. Categorical perception of melodic tone

1. Melodic tone discrimination

For the melodic tone AX discrimination (three-note arpeggio varying in D note, e.g., [CX_1E, CX_2E]), a mixed two-way ANOVA with Group as the between-subject factor and Melodic pair (step) (Pair1–2, Pair2–3, Pair3–4,…, Pair17–18) as repeated measures was conducted. The measurement was discrimination accuracy ranging from 0 to 1, with a score of 1 meaning 100% correct discrimination of the music tone pairs, calculated in the same way as that for the Mandarin tone discrimination.

A significant main effect of Group was obtained [$F(2, 866) = 71.6$, $p < 0.0001$]. *Post hoc* Tukey HSD p-adjusted pair-wise comparisons indicated that EMs ($M = 0.76$) performed significantly better than ENs ($M = 0.67$, $p < 0.0001$), which in turn was better than MNs ($M = 0.60$, $p < 0.0001$). There was no significant main effect of Pair [$F(16, 866) = 0.3$, $p = 1$], indicating that the performance across the steps was comparable, with no significant discrimination peak. Additionally, no interaction between Group $\times$ Pair was observed [$F(32, 866) = 0.3$, $p = 1$]. Figure 7 displays the

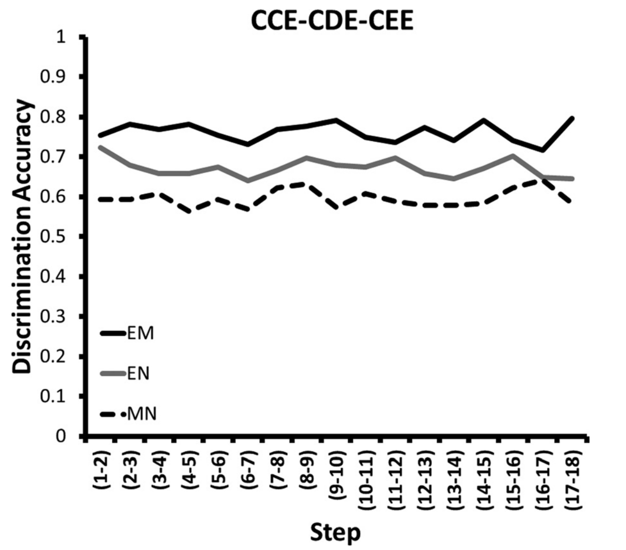

FIG. 7. Melodic tone discrimination curves for each group (EM, English musicians; EN, English non-musicians; MN, Mandarin non-musicians). The *y* axis shows the discrimination scores ranging from 0 to 1, with 1 being perfect (100% correct) discrimination of the melodic pairs and 0.5 being chance level. The *x* axis shows the discrimination pairs along the melodic continuum from CCE to CDE to CEE.

melodic tone discrimination scores for each of the pairs for the three groups.

2. Melodic tone identification

For the identification task, the melodic tone continua were separated as CCE-to-CDE and CDE-to-CEE. The

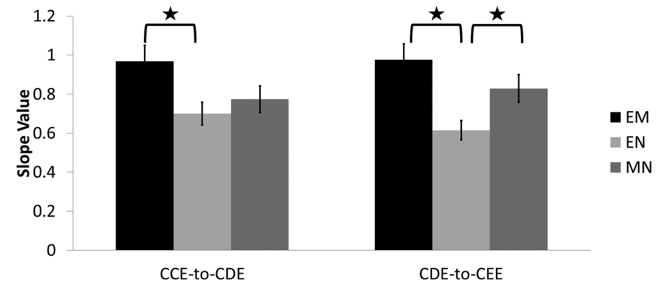

FIG. 9. Melodic tone slope (category boundary sharpness) comparisons from the logistic identification functions for each continuum (CCE-CDE, CDE-CEE) and each group (EM, English musicians; EN, English non-musicians; MN, Mandarin non-musicians). Error bars indicate standard error.

melodic tone identification results in terms of the category boundary sharpness (slope) and boundary location were derived from the generalized linear model analysis, conducted in the same way as that for the Mandarin tone identification. The category boundary location ranges from 1 to 9 for the CCE-CDE continuum (Step 1, CCE to Step 9, CDE) and from 1 to 10 for the CDE-CEE continuum (Step 9, CDE to Step 18, CEE). The melodic tone logistic identification functions as well as the corresponding discrimination curves are displayed in Fig. 8.

To compare the category boundary sharpness (slope) among groups, a one-way ANOVA using Group (EMs, ENs, MNs) as the between-subject factor was conducted for each melodic continuum. A significant Group effect was observed for the CCE-CDE continuum [$F(2, 48) = 2.5$, $p = 0.006$]. *Post hoc* Tukey HSD *p*-adjusted pair-wise group comparisons revealed a sharper category boundary for EMs than

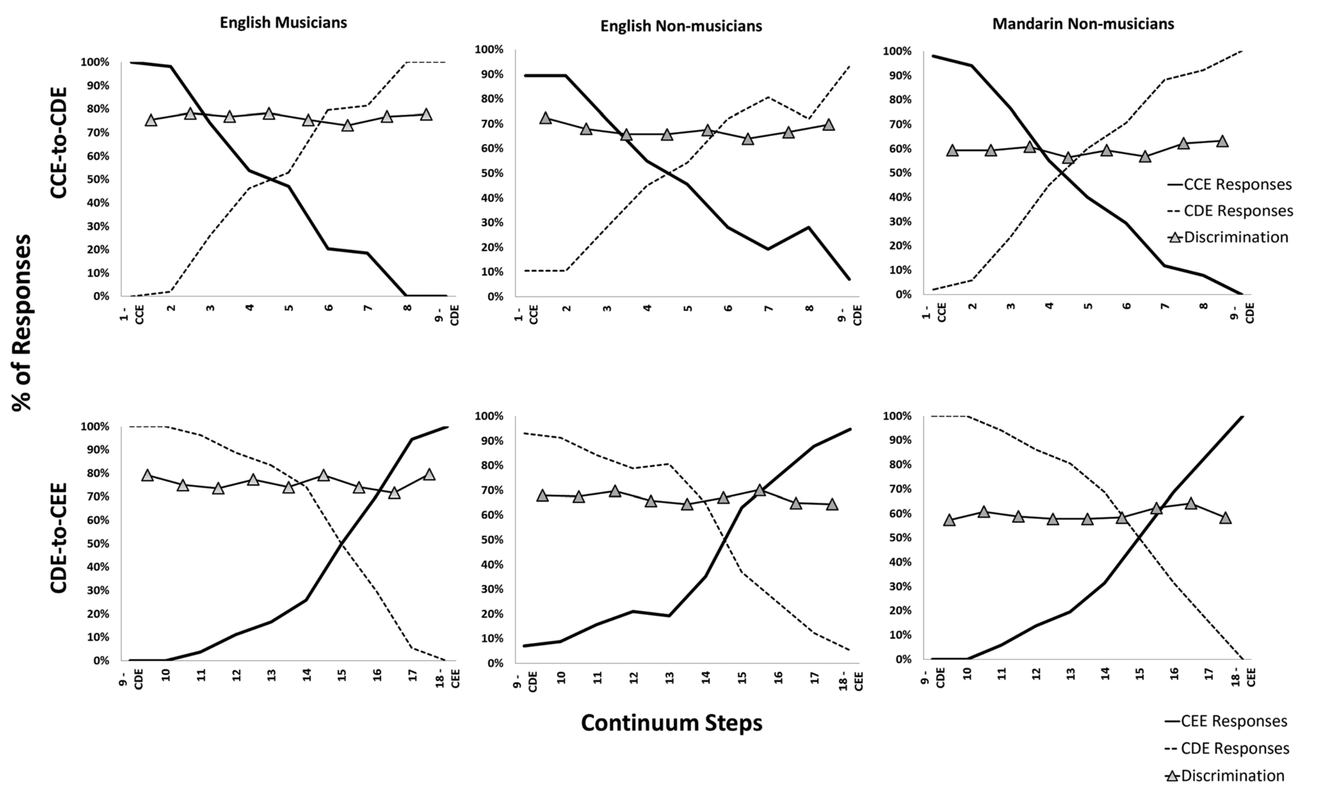

FIG. 8. Melodic tone logistic identification functions and discrimination curves for each type of continua (CCE-CDE, CDE-CEE) and each group [English musicians (EMs), English non-musicians (ENs), Mandarin non-musicians (MNs)]. The identification curves plot the percent responses of each of the melodic tone stimuli as a function of step along each of the melodic continua. CCE-CDE ranges from steps 1–9, and CDE-CEE ranges from steps 9–18, where 9 represents CDE.

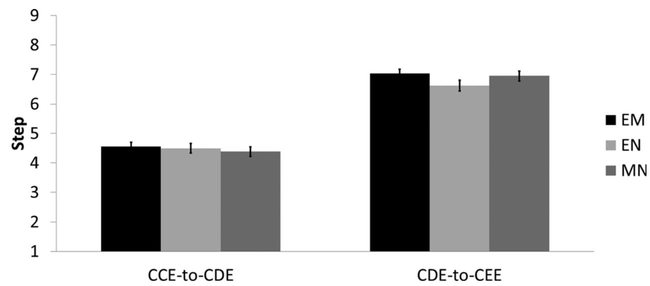

FIG. 10. Melodic tone category boundary location (step at 50% cross-over) comparisons from the logistic identification functions for each continuum (CCE-CDE, CDE-CEE) and each group (EM, English musicians; EN, English non-musicians; MN, Mandarin non-musicians). Error bars indicate standard error.

ENs[4] ($p = 0.006$), while there was no significant difference between MNs and either EMs ($p = 0.051$) or ENs ($p = 0.809$). A significant effect of Group was also obtained for the CDE-CEE continuum [$F(2, 48) = 5.0$, $p = 0.002$], with *post hoc* Tukey HSD *p*-adjusted pair-wise comparisons indicating that the category boundary was significantly sharper for EMs than for ENs ($p = 0.009$). Moreover, the MNs' category boundary was also significantly sharper than ENs' ($p = 0.004$). There was no significant difference between EMs and MNs ($p = 0.912$). Figure 9 displays the group comparisons of the slope values for the melodic tone identification.

To compare the category boundary location, a one-way ANOVA with Group (EMs, ENs, MNs) as the between-subject factor was conducted for each melodic continuum. No significant main effect of Group was obtained for either continuum [CCE-CDE: $F(2, 48) = 0.20$, $p = 0.823$; CDE-CEE: $F(2, 48) = 0.7$, $p = 0.488$]. For the 9-step CCE-CDE continuum, the average boundary location was between Step 4 and Step 5 ($M = 4.5$), and for the 10-step CDE-CEE continuum, at the average boundary location was between Step 6 and Step 7 ($M = 6.9$), with the responses for both continua being in favor of CDE. Figure 10 displays the boundary locations in the melodic tone identification task.

C. Summary

The results show that for Mandarin tone discrimination across steps, EMs outperformed MNs (significantly for the falling-level stimuli and marginally for the rising-level stimuli), but ENs did not differ from either of these two groups. For Mandarin tone identification, the EMs approximated the MN patterns, showing a significantly sharper category boundary (more categorical perception) than did the ENs. For melodic tone discrimination across steps and continua, the EMs outperformed the ENs, followed by the MNs. For melodic tone identification, the MNs followed the EM patterns with a sharper category boundary relative to the ENs for the CDE-CEE stimuli, although the MNs did not differ from the ENs for the CCE-CDE stimuli. These results as well as the original predictions are summarized in Table I and discussed in terms of the bi-directional influence of musical and linguistic experience on Mandarin and melodic tone perception, respectively.

IV. DISCUSSION

A. Effects of musical training on categorical perception of lexical tones

The first question of the present research addresses the extent to which musical training affects categorical perception of lexical tones. The results show EMs' superiority over MNs in discriminating Mandarin tonal exemplars. Specifically, the finding that the EMs had better discrimination scores across all pairs than the MNs but the two groups had comparable scores at category boundary (discrimination peak location) indicates that the EMs outperformed the MNs in discriminating $F0$ differences within tonal categories. As predicted, the Mandarin natives' performance is consistent with the previous findings that their experience with categorical tonal distinctions decreases their sensitivity to within-category $F0$ differences (Hallé *et al.*, 2004; Peng *et al.*, 2010). The results that the ENs' discrimination was on par with the EMs' appears to counter the prediction of musicians' advantage over non-musicians' in tonal discrimination due to musical training experience. However, the current patterns find some support from Chandrasekaran *et al.* (2009) in that, although English musicians showed greater MMNs than non-musicians at the pre-attentive stage, both groups were equally more accurate than Mandarin non-musicians in discriminating within-category $F0$ differences of Mandarin tonal distinctions at the attentive stage. Thus, despite the fact that musical training experience may enhance sensitivity to pitch, this ability may not transfer to enhance discrimination of speech-related $F0$ differences, as the latter demands listeners to ignore such irrelevant, minute physical distinctions in order to make categorical judgements.

These findings indicate that the facilitative effects of musical training experience on linguistic tone perception may occur at a higher categorical level, as musical training also focuses on extraction of abstract rules and regularities to form

TABLE I. Summary of the predictions and results for Mandarin and melodic tone perception by English musicians (EM), English non-musicians (EN), and Mandarin non-musicians (MN). ">": more accurate discrimination or identification ($p < 0.05$); "≥": more accurate or no difference; "=": no difference ($p > 0.05$); "*": marginal difference for rising-level ($p = 0.057$), significant for falling-level ($p < 0.05$). The results that are not consistent with the predictions are denoted in boldface.

	Discrimination		Identification	
	Prediction	Result	Prediction	Result
Mandarin tone	EM > EN ≥ MN	EM > MN*, **EM = EN**, MN = EN	MN ≥ EM > EN	MN > EM > EN (rising-level) MN = EM > EN (falling-level)
Melodic tone	EM > EN ≥ MN	EM > EN > MN	EM ≥ MN > EN	EM = MN, **MN = EN**, EM > EN (CCE-CDE) EM = MN > EN (CDE-CEE)

discrete categories in music. The results from Mandarin tone identification indeed show that the EMs patterned similarly with the native MNs with more categorical perception than the ENs. The EMs' greater sensitivity to categorical distinctions of the tones relative to the ENs' suggests that their experience with classifying $F0$ differences into categories gained from musical training has been transferred to linguistic tonal categorization. These findings are consistent with the prediction of this study, providing empirical evidence to support the theoretical account that domain-specific (musical training) experience with categories enhances domain-general category learning mechanisms and subsequently facilitates cross-domain categorization (Asaridou and McQueen, 2013; Kraus and Chandrasekaran, 2010; Patel, 2011).

In addition to the domain-general positive transfer, the present study has also revealed domain-specific patterns where musicians' specific musical training experience did not benefit their linguistic tone categorization. In the identification of the Mandarin rising-level tonal exemplars, the EMs' category boundary was less steep as compared to the Mandarin natives, showing that the EMs did not reach the native-level sensitivity to the Mandarin rising and level tone distinctions. Furthermore, in identifying the Mandarin falling-level tonal exemplars, the EMs had a tendency to favor the falling over level tones, to a greater extent as compared to the Mandarin natives. The EMs' musical pitch training experience may account for these categorization patterns that are not "native-like." Particularly, compared to the Mandarin natives, the EMs' lack of sensitivity between contour (rising) and level tones as well as their preference of contour to level tones may be attributed to the musicians' greater familiarity with directional than level tones, as melodies in music prevalently involve changes in pitch direction (Peretz and Coltheart, 2003; Schön *et al.*, 2004). These results are consistent with the previous findings that musical training experience enhances the sensitivity to $F0$ contour (rather than $F0$ height) in linguistic tones (Francis *et al.*, 2008; Huang and Johnson, 2011; Lee *et al.*, 2014).

Together, the music-to-speech results suggest that the facilitative effects of musical training experience on Mandarin tone perception occur at the categorical level more than at the sensory-acoustic level. Furthermore, the effects of musical experience may not be exclusively facilitative, given that musicians' enhanced sensitivity to within-category $F0$ distinctions may not enhance categorical-level linguistic tone distinctions, and that their enhanced categorization ability in music may not aid the perception of linguistic tones with different categorical features.

B. Effects of tone language experience on melodic tone perception

With respect to the overall accuracy for the melodic tone discrimination task, the MNs performed more poorly than the ENs and EMs. Whereas the musicians' superiority over the non-musicians is expected and consistent with previous findings (e.g., Hutka *et al.*, 2015), the MNs' failure to outperform the ENs appears to contradict some previous research predicting and showing that tonal $L1$ non-musicians benefit from their linguistic tone background in detecting melodic changes (Bidelman *et al.*, 2013; Stevens *et al.*, 2011). However, as reviewed earlier, further findings also reveal that there may not be an $L1$ advantage in melodic discrimination when the $F0$ variations are smaller than those occurring between $L1$ tone categories (Bidelman *et al.*, 2013; Hutka *et al.*, 2015). The current results of Mandarin natives' poor discrimination of melodic variations support these findings. It appears that the Mandarin listeners' $L1$ tone experience has interfered with their discrimination of small (4 Hz) $F0$ changes in melodies, which are smaller than Mandarin categorical tonal distinctions (Peng *et al.*, 2010; Xu *et al.*, 2006). This finding is consistent with the previous claim that tonal-$L1$ listeners' experience with classifying pitch into categories decreases their sensitivity to subtle $F0$ differences within a speech tone category (Hallé *et al.*, 2004; Peng *et al.*, 2010), and has further extended this claim to cross-domain discrimination of melodic changes. Moreover, the challenge that the Mandarin listeners face with regard to fine-grained pitch discrimination may also be attributed to the mismatch of specific melodic tonal features in relation to tone language listeners' $L1$ tonal features (Bent *et al.*, 2006; Xu *et al.*, 2006). Interference occurs when the features of melodic tones do not match the features of linguistic tones, as MNs' mentally-stored lexical tones had to compete against the perceived melodic tones during the discrimination task (Alexander *et al.*, 2011; Peretz *et al.*, 2011). For instance, the present melodic tonal discrimination involves the detection of the acoustic changes of an intermediate note in the context of notes C and E, which essentially requires detection of subtle changes in $F0$ height of a single note. This task presumably caused difficulty for the Mandarin listeners, as Mandarin only contains one level tone. A similar challenge has been revealed within the linguistic domain, in that native Mandarin listeners had difficulty distinguishing the three level tones in Cantonese (Qin and Mok, 2012). Thus, it is possible that detecting the level pitch differences in music was difficult for the native Mandarin listeners as they typically rely on $F0$ contour rather than $F0$ height as the primary cue to tone perception (Huang and Johnson, 2011; Lee *et al.*, 2014).

Such experience-dependent influence is also reflected at the categorical level. The identification results reveal that the MNs' melodic tonal perception category boundary was sharper than the ENs' in the CDE-CEE continuum, indicating that MNs perceive melodic tones in a more categorical manner. However, there was no difference between the MN and EN groups in the identification of the CCE-CDE continuum. Mandarin listeners' experience with the specific categorical pitch patterns of Mandarin tone may have affected how they perceive the melodic differences. Particularly, a continuum from CDE to CEE involves a significant increase in $F0$ at the 1/3 point in time of the melodic string, followed by stepwise $F0$ increases in the second 1/3 of the stimulus towards equalizing the last 1/3 with the highest note. This pattern is similar to a rising tone contour in Mandarin, where the onset of $F0$ rise (i.e., the turning point) occurs at about 1/3 into the tone (see Fig. 1). Thus, for the MNs, to identify tones along the CDE-CEE continuum is similar to distinguishing two categories from a non-typical rising tone (at the CDE end) to a typical rising tone (at the CEE end). In contrast, the CCE-CDE continuum mostly involves drastic tone changes from the last 1/3 of the melodic string, which does not have counterparts in Mandarin. In this latter case, the MNs, just as the ENs, might have relied more on acoustic differences resulting in less categorical perception of the melodies. Thus, MNs' linguistic tone experience has strengthened the categorization of the CDE-CEE continuum, consistent with the current as well as the theoretical prediction that experience with linguistic tone categorization can be transferred across domains to facilitate melodic categorization (Asaridou and McQueen, 2013). In contrast, such advantage is absent if tonal-$L1$ listeners' $L1$ tonal features are incompatible with melodic tonal features (Bidelman *et al.*, 2013; Peretz *et al.*, 2011), as shown from the CCE-CDE results.

To summarize, the speech-to-music results indicate that the positive effects of linguistic tone experience on melodic tone perception are modulated at the categorical level, where experience with tonal categorization in speech enhances sensitivity to melodic categorization through domain-general mechanisms common to both speech and music. However, both discrimination and identification data show that domain-specific speech features incompatible with melodic tonal features also interfere with melodic tone perception, both in the discrimination and identification processes.

C. Bi-directional influence of music and speech on tone perception

Together, the music-to-speech and speech-to-music results reveal consistent patterns of bi-directional transfer effects, incorporating both domain-general and domain-specific processes. First, musical experience and tone language experience are shown to be mutually transferable to facilitate cross-domain tone categorization but not discrimination of fine-grained tonal changes, suggesting that domain-general influence primarily takes effect at the abstract, categorical level.

In terms of domain-specific processes, experience with both language-specific and melodic-specific tonal features appear to interfere with transfer effects, where benefits depend on the compatibility of tonal features in the two domains.

The bi-directional transfer of tonal categorization supports our theoretical prediction that domain-specific experience with categories may enhance domain-general category learning mechanisms which in turn facilitate cross-domain categorization (Asaridou and McQueen, 2013; Kraus and Chandrasekaran, 2010; Patel, 2011). These patterns suggest that experience which leads to increased efficiency of sound category learning could benefit the categorization in both music and speech under domain-general processes (Asaridou and McQueen, 2013). On the other hand, these cross-domain effects are also determined by domain-specific processes, in that specific experience with speech or music tones may also interfere with cross-domain categorization patterns (Bidelman *et al.*, 2013; Peretz *et al.*, 2011). For example, the current music-to-speech results reveal experience-dependent effects, where the EMs' categorization bias in favor of Mandarin contour tones may be attributed to their musical training experience with melodic tones which are directional in nature (Peretz and Coltheart, 2003; Schön *et al.*, 2004). Similarly, the speech-to-music results show that benefit of tone language experience depends on the compatibility of $L1$-specific tone features and the target melodic tone features, with enhanced categorization ability for the melodic patterns that have similar $L1$ tonal counterparts. These patterns are in line with the theoretical proposition that domain-specific experience with pitch categorization can benefit categorization in other domains on condition that the acoustic features characterizing the target categories are compatible with those from listeners' specific experience (Asaridou and McQueen, 2013; Bidelman *et al.*, 2013; Hutka *et al.*, 2015).

The discrimination results from both music-to-speech and speech-to-music directions also reveal converging evidence that the facilitative effects between speech and music tonal transfer did not occur at the sensory-acoustic level. This lack of transfer is in line with the previous claim that the encoding of such domain-general information to be used in cross-domain processing is affected by domain-specific experience (Chandrasekaran *et al.*, 2009; Patel, 2011). Thus, although musical experience with high precision pitch detection enhances general pitch acuity, this enhanced ability may not be relevant to the perception of speech-related $F0$ variations, as the latter demands listeners' to ignore minor acoustic differences in order to make categorical judgements. This claim is further supported by the speech-to-music results, showing that the MNs' experience with tones in their $L1$ did not facilitate their discrimination of melodic tonal variations, presumably because their long-term linguistic tone experience has tuned them to broad categorical distinctions while normalizing for within-category acoustic variations (Bidelman *et al.*, 2013; Hallé *et al.*, 2004; Peng *et al.*, 2010).

V. CONCLUDING REMARKS

The current results show bi-directional influence of pitch experience between speech and music tones in a single design, thus providing new data supporting the speech-music relationship theories which have been developed mainly based on uni-directional evidence (Asaridou and McQueen, 2013; Patel, 2008, 2011). Moreover, the current data reveal that cross-domain transfer of pitch proficiency occurs at the abstract, categorical level, offering support for the shared sound category learning mechanism hypothesis in speech contexts which had not been tested before (Asaridou and McQueen, 2013; Patel, 2008). Furthermore, the current findings indicate that the nature and extent of benefits from cross-domain resources are experience-driven, depending on the compatibility between previous pitch experience and the demands of target pitch perception. Together, these patterns suggest shared sound category learning mechanisms governing tonal categorization across cognitive domains, which are modulated by long-term pitch exposure through musical or linguistic experience, clearly demonstrating the integration of domain-general and domain-specific mechanisms underlying pitch processing (Zatorre and Gandour, 2008).

The current findings also have practical implications. For example, previous research has proposed various merits of music-based training to benefit second language (*L*2) learning or language restoration in a clinical setting (Asaridou and McQueen, 2013; Cooper and Wang, 2012; Patel, 2011). Compared to speech-based training, music-based training allows *L*2 learners of a tone language to focus their attention on an acoustic signal without having to extract conceptual meaning from the signal; and for those individuals with language deficits due to brain damage, musical training may recruit alternative cortical resources for tonal restoration in speech. However, results from the current study further suggest that not all types of musical experience would facilitate speech pitch learning. For instance, given the results that positive music-to-speech transfer occurs at the categorical level, musical training that focuses on enhancing categorization ability should benefit speech tone categorization. However, although musical training aiming at enhancing pitch acuity may result in increased sensitivity to fine-grained *F*0 differences, it may not benefit speech tone learning, since the latter requires listeners to ignore such subtle differences in order to make categorical distinctions. Likewise, the current results show a compatibility-driven influence, indicating that only those melodic patterns compatible with language-specific tonal patterns could benefit speech tone learning, whereas incompatible patterns may restrict or even inhibit positive transfer. These patterns suggest the need for targeted training algorisms, in keeping with the OPERA-based prediction that the benefits of musical training depend on specific features emphasized in training (Patel, 2011). Thus, musical training with the intention of facilitating pitch processing in speech should focus on those features that overlap with speech pitch in a particular language.

ACKNOWLEDGMENTS

This study was funded by a research grant from the Natural Sciences and Engineering Research Council of Canada to Y.W. (NSERC Discovery Grant No. 312457-2011). We would like to thank Roxane Chan, Ashley Knight, Courtney Lawrence, Keith Leung, Joy Zhao, and other members from the Language and Brain Lab at Simon Fraser University for their assistance in stimulus development, recruiting, and data collection. We also thank Dr. Allard Jongman, Dr. Joan Sereno, Dr. Yang Zhang, as well as Dr. Tessa Bent and the two anonymous reviewers for their valuable comments and suggestions at various stages of this project. Special thanks to Ian Bercovitz from the Department of Statistics at Simon Fraser University for assistance in statistical analysis. Portions of this research were conducted as D.C.'s master's thesis.

[1]The two types of tone continuum, rising-level and falling-level, are analyzed separately, because these continua were developed by manipulating different acoustic tonal attributes. Data from the two melodic continua are presented separately due to similar reasons.

[2]As noted in Sec. II, Direction refers to the direction from which a tone continuum was created, e.g., level-to-rising was created based on a natural level tone, while rising-to-level was created based on a natural rising tone. This was intended to be a counterbalancing measure to avoid any potential perception bias among the three listener groups in terms of how a tone continuum was created. This factor was not the focus of this study. Thus only significant interactions of Direction and Listener Group would be relevant and reported in this section. (In fact, none of the results showed any significant interactions, indicating that the observed group differences were not affected by how a tone continuum was created.)

[3]Two outliers have been excluded from the EM group in this analysis due to their extremely high values (i.e., greater than two SDs as detected by the statistical analysis software, JMP, SAS Institute).

[4]Three outliers have been excluded for the CCE-CDE continuum and one outlier has been excluded for the CDE-CEE continuum from the EN group in this analysis due to their extremely high values (i.e., greater than two SDs as detected by the statistical analysis software, JMP, SAS Institute).

Alexander, J. A., Bradlow, A. R., Ashley, R. D., and Wong, P. (**2011**). "Music-melody perception in tone-language and non-tone-language speakers," in *Proceedings of the Psycholinguistic Representation of Tone Conference*, Hong Kong, pp. 1–4.

Asaridou, S. S., and McQueen, J. M. (**2013**). "Speech and music shape the listening brain: Evidence for shared domain-general mechanisms," Frontiers Psychol. **4**, 1–14.

Bent, T., Bradlow, A. R., and Wright, B. A. (**2006**). "The influence of linguistic experience on the cognitive processing of pitch in speech and non-speech sounds," J. Exp. Psychol.: Human Percept. Perform. **32**, 97–103.

Besson, M., Chobert, J., and Marie, C. (**2011**). "Language and music in the musician brain," Lang. Linguistics Compass **5**, 617–634.

Bidelman, G., Hutka, S., and Moreno, S. (**2013**). "Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: Evidence for bidirectionality between the domains of language and music," PloS One **8**, e60676.

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. Cognit. Neurosci. **23**, 425–434.

Chandrasekaran, B., Krishnan, A., and Gandour, J. T. (**2009**). "Relative influence of musical and linguistic experience on early cortical processing of pitch contours," Brain Lang. **108**, 1–9.

Cooper, A., and Wang, Y. (**2012**). "The influence of linguistic and musical experience on Cantonese word learning," J. Acoust. Soc. Am. **131**, 4756–4769.

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?," Cognitive Process. **7**, 203–207.

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.

Deutsch, D., Henthorn, T., Marvin, E., and Xu, H. S. (**2006**). "Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period," J. Acoust. Soc. Am. **119**, 719–722.

Francis, A. L., Ciocca, V., Ma, L., and Fenn, K. (**2008**). "Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers," J. Phonetics **36**, 268–294.

Galvin, J. J., III, Fu, Q. J., and Nogaki, G. (**2007**). "Melodic contour identification by cochlear implant listeners," Ear Hear. **28**, 302–319.

Gerrits, E., and Schouten, E. H. (**2004**). "Categorical perception depends on the discrimination task," Percept. Psychophys. **66**, 363–376.

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. Phonetics **32**, 395–421.

Harrison, P. (**1996**). "An experiment with tone," UCL Working Papers in Linguistics **8**, 575–593.

Huang, T., and Johnson, K. (**2011**). "Language specificity in speech perception: Perception of Mandarin tones by native and nonnative listeners," Phonetica **67**, 243–267.

Hutka, S., Bidelman, G. M., and Moreno, S. (**2015**). "Pitch expertise is not created equal: Crossdomain effects of musicianship and tone language experience on neural and behavioural discrimination of speech and music," Neuropsychologia **71**, 52–63.

Koelsch, S., Schröger, E., and Tervaniemi, M. (**1999**). "Superior pre-attentive auditory processing in musicians," Neuroreport **10**, 1309–1313.

Kraus, N., and Chandrasekaran, B. (**2010**). "Music training for the development of auditory skills," Nature Rev. Neurosci. **11**, 599–605.

Lee, C. Y., Lekich, A., and Zhang, Y. (**2014**). "Perception of pitch height in lexical and musical tones by English-speaking musicians and nonmusicians," J. Acoust. Soc. Am. **135**, 1607–1615.

Liu, C. (**2013**). "Just noticeable difference of tone pitch contour change for English- and Chinese-native listeners," J. Acoust. Soc. Am. **134**, 3011–3020.

Marie, C., Delogu, F., Lampis, G., Belardinelli, M. O., and Besson, M. (**2011**). "Influence of musical expertise on segmental and tonal processing in Mandarin Chinese," J. Cognitive Neurosci. **23**, 2701–2715.

Moore, C. B., and Jongman, A. (**1997**). "Speaker normalization in the perception of Mandarin Chinese tones," J. Acoust. Soc. Am. **102**, 1864–1877.

Patel, A. D. (**2008**). *Music, Language, and the Brain* (Oxford University Press, New York, NY).

Patel, A. D. (**2011**). "Why would musical training benefit the neural encoding of speech? The OPERA hypothesis," Frontiers Psychol. **142**, 1–14.

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., 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. Phonetics **38**, 616–624.

Peretz, I., and Coltheart, M. (**2003**). "Modularity of music processing," Nature Neurosci. **6**, 688–691.

Peretz, I., Nguyen, S., and Cummings, S. (**2011**). "Tone language fluency impairs pitch discrimination," Frontiers Psychol. **2**, 1–5.

Pfordresher, P. Q., and Brown, S. (**2009**). "Enhanced production and perception of musical pitch in tone language speakers," Attn., Percept., Psychophys. **71**, 1385–1398.

Pisoni, D. B. (**1973**). "Auditory and phonetic memory codes in the discrimination of consonants and vowels," Attn., Percept., Psychophys. **13**, 253–260.

Qin, Z., and Mok, P. (**2012**). "The perception of speech and non-speech tones by tone and nontone language listeners," in *Proceedings of Speech Prosody 2012*, Shanghai, pp. 366–369.

Schön, D., Magne, C., and Besson, M. (**2004**). "The music of speech: Music training facilitates pitch processing in both music and language," Psychophysiol. **41**, 341–349.

So, C. K., and Best, C. T. (**2010**). "Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences," Lang. Speech **53**, 273–293.

So, C. K., and Best, C. T. (**2011**). "Categorizing Mandarin tones into listeners' native prosodic categories," Poznań Studies Contemp. Linguist. **47**, 133–145.

Stevens, C. J., Keller, P. E., and Tyler, M. D. (**2011**). "Tonal language background and detecting pitch contour in spoken and musical items," Psychol. Music **41**, 59–74.

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.

Xu, Y. (**1994**). "Production and perception of coarticulated tones," J. Acoust. Soc. Am. **95**, 2240–2253.

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.

Yang, C., and Chan, M. K. M. (**2010**). "The perception of Mandarin Chinese tones and intonation by American learners," J. Chinese Lang. Teachers Assoc. **45**, 7–36.

Zatorre, R. J., and Gandour, J. T. (**2008**). "Neural specializations for speech and pitch: Moving beyond the dichotomies," Philos. Trans. R. Soc. B **363**, 1087–1104.

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. Cognitive Process. **27**, 184–209.