Effect of musical experience on learning lexical tone categories

T. Christina Zhao[a] and Patricia K. Kuhl
Institute for Learning & Brain Sciences, University of Washington, Seattle, Washington 98195

(Received 2 April 2014; revised 18 January 2015; accepted 11 February 2015)

Previous studies suggest that musicians show an advantage in processing and encoding foreign-language lexical tones. The current experiments examined whether musical experience influences the perceptual learning of lexical tone categories. Experiment I examined whether musicians with no prior experience of tonal languages differed from nonmusicians in the perception of a lexical tone continuum. Experiment II examined whether short-term perceptual training on lexical tones altered the perception of the lexical tone continuum differentially in English-speaking musicians and nonmusicians. Results suggested that (a) musicians exhibited higher sensitivity overall to tonal changes, but perceived the lexical tone continuum in a manner similar to nonmusicians (continuously), in contrast to native Mandarin speakers (categorically); and (b) short-term perceptual training altered perception; however, there were no significant differences between the effects of training on musicians and nonmusicians. © *2015 Acoustical Society of America*.
[http://dx.doi.org/10.1121/1.4913457]

[PBN] Pages: 1452–1463

I. INTRODUCTION

Comparing music and speech processing has been of strong interest to theorists for some time. Studies have shown common as well as differential neural correlates for music and speech processing (Zatorre and Gandour, 2008), and reciprocal influences between the two domains (Deutsch *et al.*, 2006; Marques *et al.*, 2007). One important acoustic feature shared by music and speech sounds is pitch. In recent years, converging evidence strongly argues for a particularly interesting relation between the two domains regarding pitch processing: Individuals with years of musical training exhibit advantages in processing pitch contours in speech (Alexander *et al.*, 2005; Chandrasekaran *et al.*, 2009; Marques *et al.*, 2007; Schon *et al.*, 2004; Wong *et al.*, 2007).

Pitch is the psychological percept of fundamental frequency. In music, the change in pitch conveys melody. In speech, pitch variations carry critical information, such as the intonation contour that indicates a statement or question, as well as the emotional state of the speaker (e.g., happy vs sad). In tonal languages, which make up over half of the world's languages, pitch contour patterns on the syllable level are contrastive; i.e., a change in pitch contour changes word meaning. These pitch contour patterns are called lexical tones (Hyman, 2001). For example, in Mandarin Chinese, there are four lexical tones with distinct pitch contour patterns: Tone 1 (hereafter T1), level; Tone 2 (hereafter T2), rising; Tone 3 (hereafter T3), falling-rising; and Tone 4 (hereafter T4), falling (Fig. 1). Therefore, lexical tone provides an excellent opportunity for examining relations between musical experience and the ability to learn non-native speech sounds, specifically lexical tones, in speakers of a non-tonal language (Patel and Iversen, 2007).

The acquisition of non-native lexical tones by speakers of a non-tonal language requires sensitivity to the acoustic differences of the sounds as well as the formation of robust phonemic categories. That is, one must assign highly acoustically variable speech sounds efficiently and consistently to invariant phonemic groups. To accomplish this, listeners must be able to ignore acoustic differences within the same phonemic group while remaining sensitive to the acoustic differences across phonemic boundaries, resulting in a "warped perceptual space" (Kuhl and Iverson, 1995). For example, the acoustic signals of the word má (T2, meaning "hemp"), produced by the same speaker at two different times, will be acoustically different. Nevertheless, a native listener identifies them to be the same word within milliseconds. This phenomenon has been extensively studied over the past 50 years, and multiple theories have been proposed to account for this behavior (Harnad, 1987; Kuhl and Iverson, 1995; Liberman *et al.*, 1957). In the original study

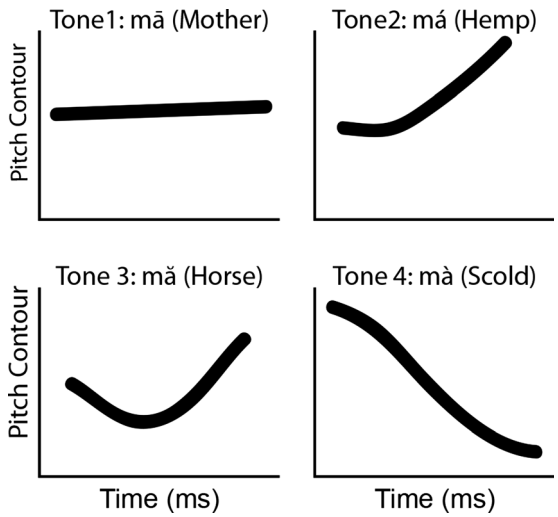

FIG. 1. Schematic representations of the four lexical tone pitch contour patterns in Mandarin Chinese, using syllable "ma" as an example.

[a]Author to whom correspondence should be addressed. Electronic mail: zhaotc@uw.edu

of English speech sounds, Liberman *et al.* (1957) created a continuum of synthesized syllables that varied continuously in one acoustic dimension (i.e., second-formant transition). Native English-speaking participants were instructed to label and discriminate the sounds on the continuum. The labeling behavior changed sharply when the sounds crossed the phonemic boundary (/ba/ vs /da/), and a sensitivity peak was observed in the discrimination task at the location of the boundary. This pattern of behavior was taken as evidence that listeners process these synthesized speech sounds in a categorical manner (Liberman *et al.*, 1957). Cross-linguistic and training studies suggest that this perceptual phenomenon can be explained by one's long-term memory of the phoneme categories, resulting from years of experience with the native language (i.e., top-down processing) (Iverson *et al.*, 2003). In the case of tonal languages, researchers have created continuums of lexical tones by gradually varying the pitch contours and testing native tonal language speakers and speakers of non-tonal languages. Native speakers of tonal languages have been shown to perceive lexical tone continua in a more categorical manner than speakers of non-tonal languages; that is, they exhibited a sharper change in the labeling identification and a more prominent sensitivity peak in the discrimination (Halle *et al.*, 2004; Peng *et al.*, 2010; Xu *et al.*, 2006). We note, however, that the musical experience of the participants was not a controlled variable in these studies.

Previous research has provided physiological and behavioral evidence that musicians exhibit advantages in processing and encoding the acoustic information/differences in lexical tones. At the brainstem level, musicians show a more faithful frequency-following response, which indicates better preservation of the original signal (Bidelman *et al.*, 2011; Wong *et al.*, 2007). At the cortical level, larger mismatch-negativity responses are elicited by deviant tones in musicians, indicating higher sensitivity to the differences between the lexical tones (Chandrasekaran *et al.*, 2009). Behaviorally, musicians identify and discriminate different lexical tones with greater accuracy even when the speech signal is degraded (Alexander *et al.*, 2005; Lee and Hung, 2008). Such evidence supports the hypothesis that enhanced frequency processing and encoding mechanisms related to musical training can be transferred to the speech-processing domain (Kraus and Chandrasekaran, 2010; Patel, 2011).

Building on the studies demonstrating the advantages that non-tonal language speaking musicians exhibit in detecting acoustic differences in lexical tones, we investigated whether extensive experience with music influences the perceptual learning process for lexical tone categories in two experiments. In Experiment I, we examined whether musicians who had no prior experience with tonal languages differed from nonmusicians in the perception of a lexical tone continuum. Monolingual English-speaking musicians and monolingual English-speaking nonmusicians completed discrimination and identification tasks that measured their perception of a tone continuum. Their performance was compared to that of Mandarin-speaking nonmusicians, in order to determine whether musical training in English speakers is associated with lexical tone perception similar to that of native Mandarin speakers. In Experiment II, we explored whether short-term perceptual training on lexical tones altered the perception of the tone continuum differentially in English-speaking musicians and nonmusicians.

II. EXPERIMENT I

In order to examine the effects of musical experience on the learning of lexical tone categories, it is important to first examine the differences in the perception of lexical tones by English-speaking musicians and English-speaking nonmusicians, and the differences between English speakers and native Mandarin speakers. Therefore, the goal of Experiment I is twofold: (1) To replicate results from previous studies demonstrating that native Mandarin speakers perceive a tone continuum more categorically than speakers of non-tonal languages (i.e., English speakers in this study), using a new T2–T3 continuum (T2–T3 hereafter); and (2) to examine the differences between English-speaking musicians and nonmusicians in the perception of the tone continuum. The T2–T3 contrast was chosen as the target contrast for two reasons: (1) These two tones are the most difficult to discriminate due to their similar lexical tone contours (Fig. 2), T2 has a rising pattern while T3 has an initial dip followed by a rising (Kiriloff, 1969; Shen and Lin, 1991); (2) therefore the categorical boundary between the two tones reflects

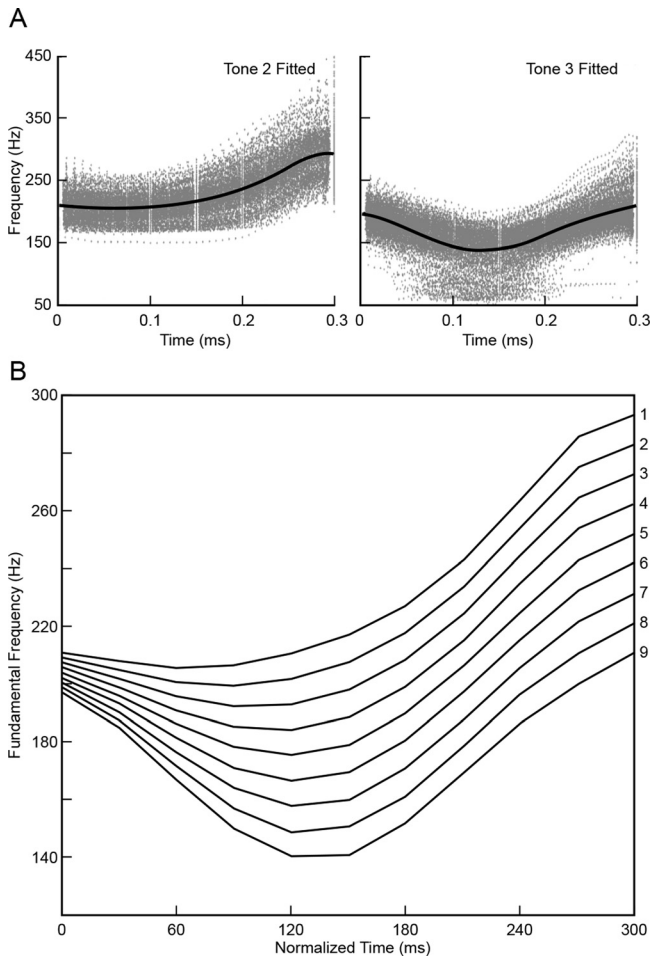

FIG. 2. (A) T2 and T3 modeled from 600 samples recorded from 5 speakers. (B) The 9-step *F*0 contour continuum from T2 (contour 1) T3 (contour 9).

linguistic categories rather than acoustic differences, which may contribute significantly to discrimination of other tone pairs. For example, discrimination of T1–T2 pair (level vs rising) can be heavily influenced by the detection of frequency modulation in the acoustic signal.

We hypothesized that English-speaking musicians and nonmusicians will perceive the tone continuum in a similar manner (continuously), in contrast to the perception of Mandarin speakers (categorically). The perception of the continuum is evaluated by a discrimination task and an identification task. We expect that in the discrimination task, Mandarin speakers will show differential sensitivities across the tone continuum while English speakers' sensitivities will remain constant. At the same time, we expect English-speaking musicians will exhibit the highest sensitivity uniformly along the continuum. For the identification task, we expect Mandarin speakers to exhibit a sharper transition (i.e., steeper slope) than English speakers.

A. Method

1. Participants

Three groups of participants were recruited: Monolingual English-speaking musicians ($n = 20$), monolingual English-speaking nonmusicians ($n = 20$), and native Mandarin-speaking nonmusicians ($n = 20$, Table I). All English-speaking musicians [10 males, mean age ± standard deviation (s.d.) = 21.80 ± 3.23] received at least 8 years of private music lessons that began before the age of 10 years (mean years of training ± s.d. = 15.34 ± 4.47); whereas all English-speaking nonmusicians (10 males, mean age ± s.d. = 21.90 ± 2.59) received less than 2 years of private music lessons that ended more than 5 years ago (mean years of training ± s.d. = 0.2 ± 0.44). Participants have no formal experience with any tonal languages, nor have they lived in tonal language speaking countries for more than 2 months. The native Mandarin-speaking nonmusicians (10 males, mean age ± s.d. = 23.90 ± 2.29) were all born and raised in Beijing, Tianjin, and Hebei areas in Mainland China and arrived in the U.S. after the age of 18 years. They have received less than 2 years of private music lessons that ended more than 5 years ago (mean years of training ± s.d. = 0.18 ± 0.37). No participant reported any history of hearing, speech, or language difficulty. All procedures were approved by the Institution Review Board and all participants were compensated for their participation.

2. Stimuli

a. Speakers. To achieve inter-speaker variability, five female native speakers of Mandarin Chinese were recruited for recording experimental stimuli. They were all born and raised in Beijing and arrived in the U.S. after the age of 18 years. All training stimuli and test stimuli were made from these recordings.

b. Material. A set of 60 syllables was selected from the original list of a previous perceptual training study (Wang *et al.*, 1999). The syllables are real words both in T2 and T3 in Mandarin Chinese and they maintain high context variability (various combinations of consonants and vowels with various syllabic structures).

c. Recordings. The speakers were instructed to read the list of syllables in both T2 and T3 with 1 s pauses between syllables to avoid influences from the previous production. Each word was produced at two speeds to achieve intra-speaker variability. Speakers were first instructed to produce the syllables at the speed of normal speech (fast speed), and then instructed to produce syllables as clearly as possible (slow speed). All recordings were completed in a sound booth using an Electrovoice 635 A microphone and a SHURE stereo headphone FP 22 amplifier (Shure Brothers, Inc., Niles, IL). Sounds were sampled at 44.1 kHz and digitized at 16 bits by the Sound Forge 6.0 software. As a result, one set of 240 stimuli (60 syllables × 2 tones × 2 speeds) were recorded from each speaker. Five sets of recordings were generated.

d. Test stimuli. A T2–T3 continuum was first created using the following method. The fundamental frequency ($F0$) contours from all the recordings (1200 in total: 240 per

TABLE I. Study design with descriptive information of subjects.

		Experiment II		
Experiment I		Perceptual training		Post-training test
English-speaking musicians		Trainees	Controls	**English-speaking musicians ($n = 20$)**
$n = 20$	⟶	**Musicians**	**Musicians**	
10 males		$n = 10$	$n = 10$	
age = 21.80 ± 3.23		5 males	5 males	
		age = 21.30 ± 3.23	age = 22.30 ± 3.34	
English-speaking nonmusicians		**Nonmusicians**	**Nonmusicians**	**English-speaking nonmusicians ($n = 20$)**
$n = 20$	⟶	$n = 10$	$n = 10$	
10 males		5 males	5 males	
age = 21.90 ± 2.59		age = 22.00 ± 2.26	age = 21.80 ± 3.01	
Native Mandarin-speaking nonmusicians				
$n = 20$				
10 males				
age = 23.90 ± 2.29				

speaker) were first extracted in Praat (Boersma and Weenink, 2009), resulting in 600 contour samples for T2 and 600 contour samples for T3. The time values of all contour samples were then scaled to 300 ms. For each tone, all 600 contour samples were first input into a database. Based on this dataset, a sixth-order polynomial function was chosen to model the contour in order to achieve the best goodness of fit without over-fitting (Zhao *et al.*, 2012) [Fig. 2(A)]. The T2 and T3 contours created from the model were used as the endpoints of the continuum. The difference (in frequency) between the two-endpoint contours was then equally divided by 8 to create 7 additional contours, yielding a 9-step T2–T3 continuum [Fig. 2(B)].

Three vowel syllables (/γ/, /u/, /y/) were chosen for test stimuli. All three syllables in both T2 and T3 are real words in Mandarin Chinese with similar word frequencies. Word frequencies were calculated by adding in all morphemes with the same syllable and tone, based on an existing public database.[1] The same five speakers were instructed to produce all three syllables in T1 (level tone). The productions were scaled to 300 ms long and low pass filtered at 5 kHz to serve as templates. For each template, the original *F*0 information was replaced with each contour from the 9-step T2–T3 continuum, using the pitch-synchronous overlap and add function in Praat. Therefore, one set of nine stimuli was generated from each template. These nine stimuli differ only in *F*0 contours, but are otherwise identical (voice quality, duration, intensity contour, etc.). In total, 15 (3 vowel syllables × 5 speakers) sets of stimuli were created as test stimuli.

3. Design and procedures

In Experiment I, all participants first completed the pitch and memory subtests of the Wing standardized test of musical intelligence (Wing, 1966). These two subtests assessed participants' ability in musical pitch discrimination and pitch memory. Then all participants completed a set of discrimination and identification tasks assessing their perception of the lexical tone continuum. For each participant, a set of test stimuli was randomly selected from the 15 sets of test stimuli described above to minimize speaker or syllable effects. All tasks were completed on a Lenovo Thinkpad T61 computer (Lenovo, Morrisville, NC) using E-prime 1.0 for stimulus presentation (Psychology Software Tools, Pittsburgh, PA). All sound stimuli were presented binaurally through a SONY dynamic stereo headphone set (SONY, Tokyo, Japan) at a comfortable intensity level.

a. Discrimination. An AX discrimination task was used. Participants were instructed to judge whether sound *X* was identical to sound A or different from A. Stimuli A and X carried *F*0 contours that were either 2 steps apart on the continuum ("different" pairs, e.g., 1-3, 3-1) or the same ("identical" pairs, e.g., 1-1, 3-3). In a single trial, a 250 ms fixation point was presented on the computer screen to signal the start of a trial. Stimuli A and X were then presented sequentially with an inter-stimulus interval (ISI) of 300 ms. Participants were required to respond on the computer keyboard within 500 ms. To eliminate response bias, the number of identical trials was adjusted be equal to different trials. This adjustment was completed by doubling the number of 5 identical pairs (3-3, 4-4, 5-5, 6-6, 7-7), because these 5 stimuli appeared in twice as many different pairs as the other 4 stimuli; for example, stimulus 3 appears in different pair 1–3 and 3–5. A total of 280 trials were presented (14 different pairs × 10 repetitions + 4 identical pairs × 10 repetitions + 5 identical pairs × 20 repetitions). All trials were presented in three blocks (see Fig. 3).

b. Identification. An AXB identification task was used. This task is modified from the classical labeling task to allow the same instructions to both native Mandarin speakers and the English speakers. Participants were instructed to judge whether sound X was more similar to sound A or sound B. The *F*0 contours of stimuli A and B remained the end points of the continuum (contour 1 and 9, see Fig. 2) while the *F*0 contour of stimulus X varied within the continuum, resulting in 18 combinations (e.g., 1-1-9, 1-2-9, 9-2-1, 9-1-1). In a single trial, a fixation point first appeared on the computer screen for 250 ms to signal the start of the trial, three sound stimuli were then presented sequentially with ISIs of 300 ms. Participants were required to respond on the computer keyboard within 500 ms. There were 180 trials in total (18 combinations × 10 repetitions). All trials were presented in 3

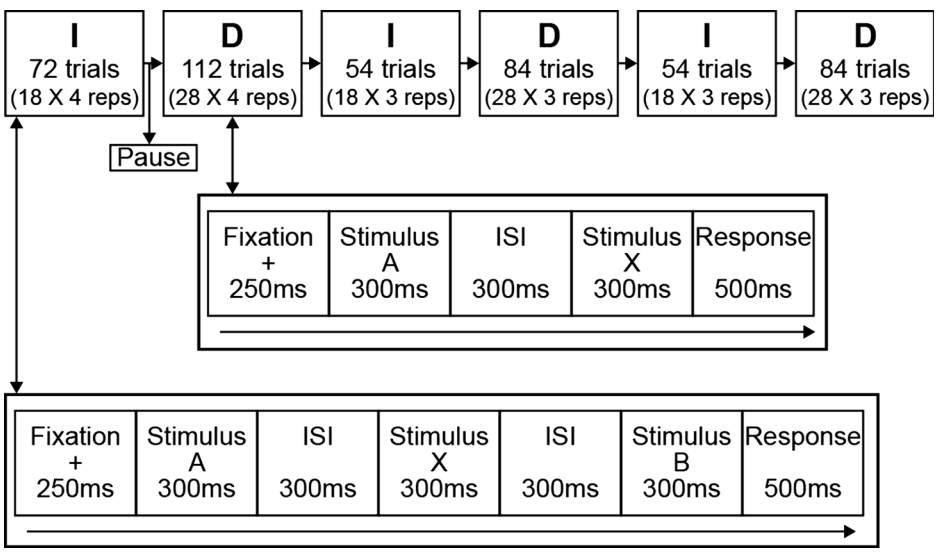

FIG. 3. The detailed procedure of the discrimination (D) and identification (I) tasks in Experiments I and II.

blocks (see Fig. 3). The six blocks of discrimination and identification tasks were alternated and participants were encouraged to take short breaks in between blocks to reduce fatigue.

B. Results

1. Data Analysis

a. Music pitch discrimination and memory. Percent correct on the pitch and memory subtests of the Wing standardized test of musical intelligence (Wing, 1966) served as measures of music pitch discrimination and music pitch memory. Both measures were obtained from all participants.

b. Perception of the tone continuum: Discrimination. The ability to reliably discriminate stimuli describes the sensitivity of each participant. From the AX discrimination data, we calculated two sensitivity measures: (1) Sensitivity for each stimulus pair (e.g., pair 1-3, 2-4, thereafter, within-pair sensitivity); and (2) sensitivity across the whole continuum (thereafter, overall sensitivity). Sensitivity was defined as the d' value calculated from percent correct ($p(c)$), Eq. (1) (Macmillan and Creelman, 2008). A response was counted as correct when the participant responded (1) different for different stimulus pairs, and (2) "same" for identical stimulus pairs,

$$d' = 2z\left(\frac{1}{2}\left\{1 + [2p(c) - 1]^{1/2}\right\}\right). \quad (1)$$

c. Perception of the tone continuum: Identification. Two measures of interest were calculated from AXB identification data: Slope and crossover. Slope indicated how sharply labeling of X changed from one end point to another as the stimulus varied along the tone continuum. Crossover marked the location of the boundary between the two endpoints, where a participant identified X as T3 (contour 9) 50% of the time.

To calculate slope and crossover, logistic regression was chosen to model each participant's response across the continuum [Eq. (2)]. In this equation, x was the stimulus number that ranged from 1 (T2 endpoint) to 9 (T3 endpoint) and p was the percent identification of x as more similar to the T3 end point. As a result, the coefficient β_1 was obtained from the model as the slope measure. Crossover was further calculated based on coefficients β_0 and β_1 [Eq. (3)]. Similar methods have been used in previous studies (Xu *et al.*, 2006),

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x, \quad (2)$$

$$\text{Crossover} = -\frac{\beta_0}{\beta_1}. \quad (3)$$

2. Perception of the tone continuum

In Experiment I, English-speaking musicians, English-speaking nonmusicians, and Mandarin-speaking nonmusicians completed the pitch and memory subtests of the Wing standardized test of musical intelligence, as well as discrimination and identification tasks that measure the perception of the tone continuum. We first validated the selection of musicians by comparing music pitch discrimination and pitch memory scores across three groups with two separate one-way analysis of variances (ANOVAs). Results revealed significant main effects for group: Pitch discrimination, $F(2,57) = 23.03$, $p < 0.001$, $\eta_p^2 = 0.45$, and pitch memory, $F(2,57) = 25.51$, $p < 0.001$, $\eta_p^2 = 0.47$. *Post hoc* tests showed that musicians scored significantly higher than both English-speaking nonmusicians and native Mandarin-speaking nonmusicians on both measures: Pitch discrimination, $p < 0.001$, and pitch memory, $p < 0.001$ (Bonferroni correction applied, Table II).

To investigate differences among groups in the perception of the tone continuum, we compared measures taken from discrimination and identification tasks for the three groups. For sensitivities measured from the discrimination task, we tested two hypotheses: (1) That Mandarin Chinese speakers exhibited differential within-pair sensitivities across the continuum while the English speakers did not, and (2) that English-speaking musicians exhibited higher sensitivities over all. We used R (R Core Team, 2012) and lme4 (Bates *et al.*, 2012) to perform a mixed effects analysis of the within-pair sensitivities across the tone continuum for the three groups (Baayen *et al.*, 2008). First of all, we constructed a full model with fixed effects of group contrasts of interest and within-pair sensitivities (along with all interactions). More specifically, the group contrasts of interest included Mandarin nonmusicians compared to all English speakers (contrast 1), English musicians compared to English nonmusicians (contrast 2), and English musicians compared to all nonmusicians (contrast 3). We also entered each subject as random effects into the model. To test the first hypothesis, we created a sub-model (1) that excluded the interaction between contrast 1 and within-pair sensitivities. Subsequently, we compared the sub-model (1) against the full model to specifically examine the effect of the interaction in question. If the interaction effect in question explains a significant amount of variance in the model, the exclusion of such term will result in significant model change. Likelihood ratio tests confirmed that the sub-model (1) significantly differed from the full model ($p = 0.048$). We further reduced the sub-model (1) to exclude interactions between contrast 2 and the tone pairs to examine whether either of English speaking groups exhibited differential sensitivities across the continuum [sub-model (2)]. Likelihood ratio tests demonstrated no significant difference ($p = 0.38$). To test the second hypothesis, we created another sub-model (3) that excluded contrast 3 as a main effect. Similarly, we compared this sub-model (3) with the full model to examine whether musicians exhibited

TABLE II. Pitch discrimination and memory scores from the three groups (Mean ± s.d.).

	English Musicians	English Nonmusicians	Mandarin Nonmusicians
Pitch Discrimination	71% ± 11%	47% ± 14%	50% ± 12%
Pitch Memory	87% ± 11%	62% ± 20%	53% ± 14%

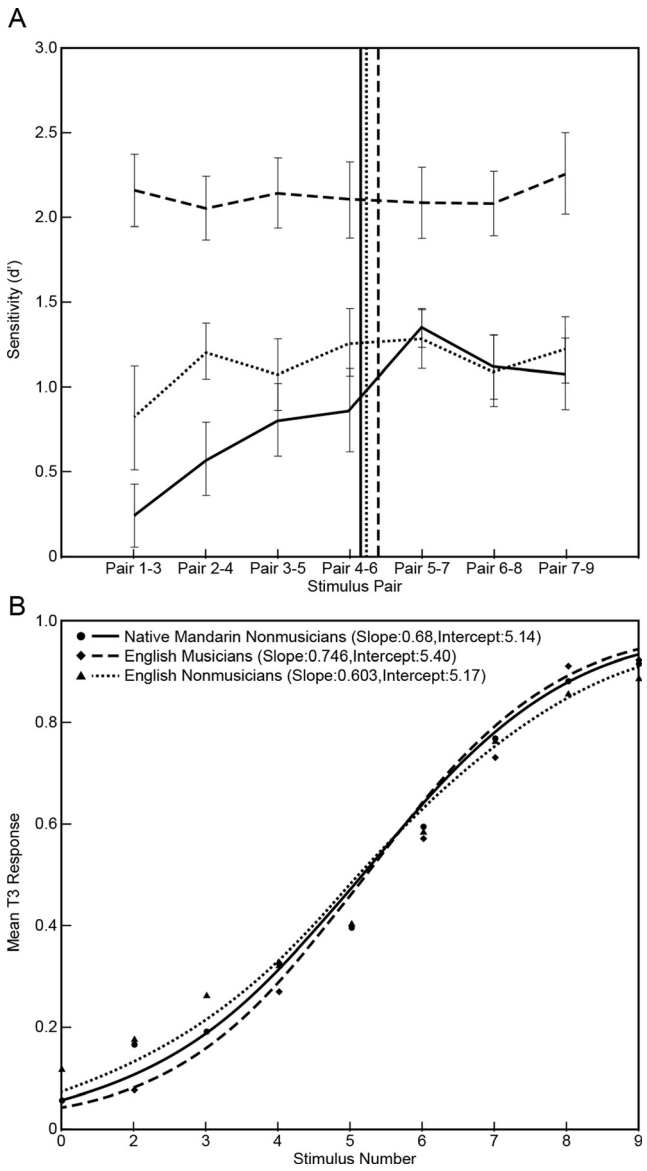

FIG. 4. (A) AX discrimination responses for the three groups in Experiment I (i.e., English-speaking musicians, English-speaking nonmusicians, and Mandarin-speaking nonmusicians). Vertical lines indicate positions of crossover points measured through AXB identification for each group. (B) AXB identification responses for the three groups in Experiment I. Each curve represents the logistic regression model with averaged slope and crossover for the corresponding group. Crossover point marks the location on the tone continuum where Mean T3 response is 0.5.

significantly different sensitivities. Likelihood ratio tested confirmed this hypothesis [$p < 0.001$, see Fig. 4(A)].

For measurements taken from the identification task, we compared the identification slope and the crossover measures across groups using two separate one-way ANOVAs. Results showed no significant main effect for group on the slope, $F(2,57) = 1.93$ ns, or the crossover measure, $F(2,57) = 0.73$ ns [see Fig. 4(B)].

We further explored how an individual's ability in music pitch discrimination and music pitch memory are associated with the perception of the tone continuum across the three groups, using multiple regression analyses. A music pitch discrimination score predicted the overall sensitivity on the tone continuum in English-speaking musicians and nonmusicians, but not native Mandarin-speaking nonmusicians, $R^2 = 0.498$, $F(5,54) = 10.727$, $p < 0.001$ [see Fig. 5(A)]; whereas it predicted identification slope in English-speaking nonmusicians, but not in English-speaking musicians or native Mandarin-speaking nonmusicians, $R^2 = 0.47$, $F(5,54) = 3.549$, $p = 0.008$ [see Fig. 5(B)]. Music pitch memory did not predict identification slope or

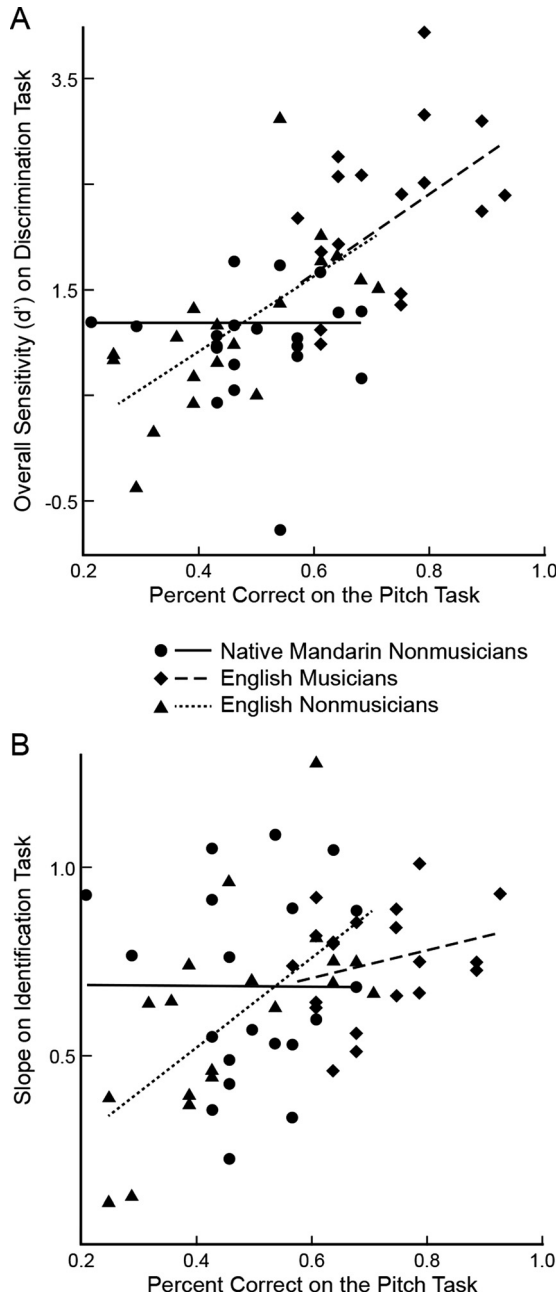

FIG. 5. (A) The relations between music pitch discrimination score and overall sensitivity on the tone continuum for the three groups in Experiment I. The fitted linear functions for each group are as follows: English-speaking musicians: $f(x) = -0.58 + 3.75x$, $p = 0.048$*; English-speaking nonmusicians: $f(x) = -0.49 + 3.57x$, $p = 0.03$*; Native Mandarin-speaking nonmusicians: $f(x) = 1.00 - 0.005x$, ns. (B) The relations between music pitch discrimination score and identification slope for the three groups in Experiment I. The fitted linear functions are as follows: English-speaking musicians: $f(x) = 0.48 + 0.37x$, ns; English-speaking nonmusicians: $f(x) = 0.04 + 1.20x$, $p = 0.028$*; Native Mandarin-speaking nonmusicians: $f(x) = 0.68 - 0.007x$, ns.

overall sensitivity in any group and was therefore excluded from further analysis.

C. Discussion

In order to examine the effects of long-term musical experience on learning lexical tone categories, it is important to first examine whether musicians' perception of lexical tones is similar to nonmusicians when neither group had prior experience with tonal languages. In Experiment I, we compared the perception of a lexical tone continuum among English-speaking musicians, English-speaking nonmusicians, and Mandarin-speaking nonmusicians using discrimination and identification tasks. We hypothesized that: (1) English-speaking musicians and English-speaking nonmusicians would exhibit a similar pattern of perception of the tone continuum, and the pattern would be different for Mandarin-speaking nonmusicians; and (2) English-speaking musicians would exhibit higher sensitivity to the acoustic differences between stimuli.

Data from the discrimination task supported our hypotheses that English-speaking musicians and English-speaking nonmusicians would exhibit a similar pattern of perception of the tone continuum, and the pattern would be different for Mandarin-speaking nonmusicians. We compared the full model and the sub-model (1) that specifically excluded the interaction terms related to Mandarin speakers and the tone pairs. Significant differences between these two models suggested that Mandarin speakers exhibited differential within-pair sensitivities across the continuum while the English speakers did not. We further compared the sub-model (1) with sub-model (2), further excluding interactions between the two English-speaking groups. No significant differences were found, suggesting the two English-speaking groups behaved similarly across the tone continuum. Furthermore the hypothesis was also supported by multiple regression analysis with musical pitch discrimination as predictor and the overall sensitivity in AX discrimination as the outcome [see Fig. 5(A)]. Musical pitch discrimination ability was significantly associated with overall sensitivity in the discrimination task in English musicians and nonmusicians, but not in native Mandarin speakers. Together, the two pieces of evidence suggested that English speakers and Mandarin speakers relied on different strategies in perceiving the tone continuum: While English speakers' ability to discriminate between stimuli relied on their sensitivities to acoustic differences, native listeners' ability to discrimination was influenced by the long term phonemic categories of the lexical tones. These results are consistent with previous research (Francis *et al*., 2003; Halle *et al*., 2004; Xu *et al*., 2006).

Data from the discrimination task also supported the hypothesis that English-speaking musicians would be more sensitive to the acoustic differences between stimuli. Results showed that English-speaking musicians exhibited a significantly higher overall sensitivity than either English-speaking nonmusicians or Mandarin-speaking nonmusicians. This result is consistent with previous research demonstrating the advantages musicians exhibited in processing and encoding foreign speech sounds (Kraus and Chandrasekaran, 2010; Patel, 2011).

One unexpected observation was the asymmetrical pattern of within-pair sensitivities in Mandarin-speaking nonmusicians: while sensitivity within T2 category (pair 1-3) was lower than sensitivity between categories (pair 5-7), sensitivity within T3 category (pair 7-9) was not. Though unexpected, this result is consistent with several recent studies investigating the characteristics of T2 and T3 (Wu, 2011; Yang, 2011). It has been speculated that T3 category in Mandarin consists of multiple important perceptual dimensions, for example, pitch contour, duration, phonation, and voice quality (Blicher *et al*., 1990; Yang, 2011). Therefore, in the current study, T3 category may be less activated in native Mandarin speakers by the pitch contour cue alone, resulting in a lesser reduction in sensitivity within the T3 category (pair 7-9).

In contrast, data from the identification task provided mixed results. First, the slope did not differ among the groups, failing to replicate results from previous research, which demonstrated sharper slopes in native speakers for other lexical tone contrasts (Halle *et al*., 2004; Xu *et al*., 2006). Second, the multiple regression analysis revealed a significant relationship between pitch discrimination and identification slope only in English-speaking nonmusicians [see Fig. 5(B)].

There are several potential explanations for these results. First, procedural differences between the current identification task and previous studies might contribute substantially to the results. The identification task employed in this study differs from the tasks in classical categorical perception studies. In a classical categorical perception study (e.g., Liberman *et al*., 1957), the identification task requires the participants to label a stimulus from a continuum between two choices (for example, to label a stimulus as /ba/ or /da/). No perceptual anchor points were provided. However, in the current study, this kind of identification task is not appropriate for English speakers without previous experience with lexical tone labels. We adapted our AXB identification approach from Halle *et al*. (2004), so that all participants were given the same anchors and can be instructed in the same manner. In a series of studies of loudness perception, Braida and Durlach proposed a model in which the performance in identification was accounted for by context coding (Braida *et al*., 1984; Durlach and Braida, 1969; Macmillan *et al*., 1988). Context coding relies on the perceptual anchors provided by either explicitly presented standards or internal landmarks. In this case, the internal landmarks (phonemic categories) in Mandarin-speaking nonmusicians may be minimized in the absence of an explicit labeling process. The resulting similar identification behaviors among all three groups may be a consequence of the explicitly presented standards. Alternatively, the native Mandarin speakers may exhibit categorical perception due to their internal phonemic categories while the English speakers are showing a similar performance due to continuous presentation of the anchors. Future research is warranted to determine identification and discrimination tasks that are

comparable in the activation level of native speakers' existing phonemic categories.

Second, working memory may play a role in AXB identification trials. Participants must first compare the acoustic differences in AX and then in XB, storing the encoded differences in their working memory to make a decision based on the context provided by A and B. Therefore, we speculate that the performance in this task not only depends on participants' sensitivity to the acoustic differences, but it also interacts with their working memory, therefore contributing to the differential regressions between the English speaking musicians and nonmusicians. Further investigation is warranted to examine how working memory influences performance on this identification task.

III. EXPERIMENT II

In Experiment II, we examined whether musical experience influences the perceptual learning of lexical tone categories. We randomly selected half of the English-speaking musicians and English-speaking nonmusicians from Experiment I to complete a perceptual training procedure for T2 and T3. The remaining participants were non-trained controls. We then compared participants' perception of the tone continuum pre-training (Experiment I) to post-training among conditions (i.e., musician trainees, musician controls, nonmusician trainees, and nonmusician controls). We were particularly interested in two questions: (1) Whether training changed the perception of the tone continuum in English speakers such that differential within-pair sensitivities emerged across the tone continuum in training groups, and (2) whether the musicians exhibited a bigger change than the nonmusicians.

A. Method

1. Participants

All English-speaking musicians ($n = 20$) and English-speaking nonmusicians ($n = 20$) from Experiment I participated in this experiment.

2. Stimuli

All 5 sets of recordings and 15 sets of test stimuli from Experiment I were used in Experiment II. The five sets of recordings were used as training stimuli in the perceptual training procedure.

3. Procedure

a. Perceptual training. Half of the English-speaking musicians and nonmusicians were randomly selected as trainees while the other half were controls (see Table I). The trainees participated in an 8-session computer-based perceptual training program completed within 2 weeks. The training material for each trainee consisted of three sets of training stimuli from speakers that were not used in the pre-test.

In each session, participants were first familiarized with examples of T2 and T3, then they completed two tasks: 180 trials of two-alternative forced choice (2AFC) and 180 trial of categorical discrimination (Wayland and Li, 2008). In a 2AFC trial, a stimulus randomly selected from the training material was presented, and the participant was instructed to indicate whether the syllable carries T2 or T3 by pressing corresponding keys on the keyboard within 1 s. In a categorical discrimination trial, two randomly selected stimuli were presented sequentially (two stimuli can vary by speaker, syllable, and tone); the participant was instructed to indicate whether the two syllables carry the same tone by pressing corresponding keys within 1 s. Feedback was provided immediately after each response. This training procedure exposed trainees to a large set of examples of T2 and T3 that were highly variable in their acoustics characteristics. Such high-variability training has been shown to benefit learners (Logan *et al.*, 1991).

b. Post-training tests. All English-speaking participants (i.e., trainees and controls, Table I) completed discrimination and identification post-training tests 2 to 3 weeks after Experiment I for the assessment of learning and generalization. Different sets of stimuli were used for each of the three post-training tests: The original pre-training (Experiment I) stimulus set, a stimulus set by the original speaker but a new vowel context (generalization 1), and a stimulus set by a novel speaker (not used in pre-test nor training), and a novel vowel context (generalization 2). The order of the three post-training tests was randomized and the procedure was identical to Experiment I. All participants took 5 to 10 min breaks between post-tests to reduce fatigue.

B. Results

1. Perceptual training

Trainees completed two types of tasks for 8 training sessions over the span of 2 weeks: 2AFC and categorical discrimination tasks. The goal of this training procedure was to provide trainees with highly variable sample sets of T2 and T3 to help them establish phoneme categories. Performance on each task was measured by percent correct. The change of performance over time for both types of tasks was evaluated by mixed two-way ANOVAs [two groups (musicians vs nonmusicians) by eight sessions (repeated measure)]. For both types of tasks, there were significant main effects of sessions [2AFC: $F(7,126) = 11.29$, $p < 0.001$, $\eta_p^2 = 0.39$, Greenhouse-Geisser corrected; categorical discrimination: $F(7,126) = 4.27$, $p = 0.011$, $\eta_p^2 = 0.78$, Greenhouse-Geisser corrected]. However, no main effect of group or group by session interactions was observed (see Fig. 6).

2. Changes in perception of the tone continuum

Changes in perception of the tone continuum from Experiment I to post-training tests were compared among four groups (English-speaking musician trainees, musician controls, English-speaking nonmusician trainees, and nonmusician controls). The discrimination and identification data was processed in the same way as Experiment I. The measures of interest were the same as in Experiment I: Within-pair sensitivities and overall sensitivities taken from discrimination, slope, and crossover taken from

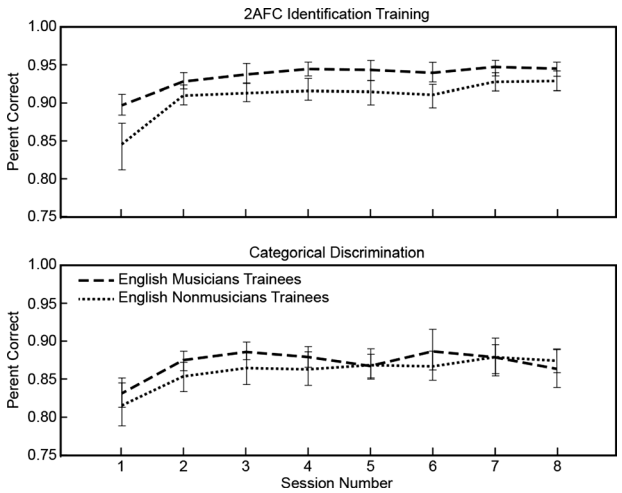

FIG. 6. Performance on 2AFC (A) and categorical discrimination (B) training tasks over the 8-session training period for musician and nonmusician trainees.

identification. Three separate comparisons were made for the three different post-training tests: Post-training test with the original stimuli vs pre-training (Experiment I), post-test with the generalization 1 stimuli (new syllable) vs pre-training, and post-test with the generalization 2 stimuli (new syllable, new speaker) vs pre-training.

a. Discrimination. We examined: (1) Whether differential within-pair sensitivities across the tone continuum emerged through perceptual training by comparing within-pair sensitivities between pre-training (Experiment I) and post-training tests, and (2) if so, whether the changes were different in musicians. Similar to Experiment I, we used R and lme4 to adopt a mixed effects modeling approach that allows us to examine specific main effects and interactions, without repeated *post hoc* tests. Procedures were identical for the analyses related to the three post-test stimuli. To test our hypotheses, we first constructed a full model (1) to include training (training vs control) and music experience (musicians vs nonmusicians) related changes (pre-training vs post-training), interacting with tone pairs. As fixed effects, we entered group contrasts of interest and within-pair sensitivities (with all interactions) into the model. More specifically, the group contrasts of interest included post-training compared to pre-training (contrast 1), training compared to control (contrast 2), and musicians compared to nonmusicians (contrast 3). Similarly, individual subjects were entered as a random effect. For hypothesis one, sub-model (1) was created to specifically exclude interaction terms between contrast 1, 2, and tone pairs. Likelihood ratio tests compared sub-model (1) against the full model. Results showed no significant difference between the two models for any of the post-test stimuli (for all three analyses, $p > 0.10$), indicating the absence of differential within-pair sensitivities related to training (see Fig. 7).

Because there is lack of differential within-pair sensitivities related to training, we did not proceed to further examine the additional interaction with music experience. We expanded the analysis by collapsing data across all levels of pairs to examine the change in overall sensitivity using

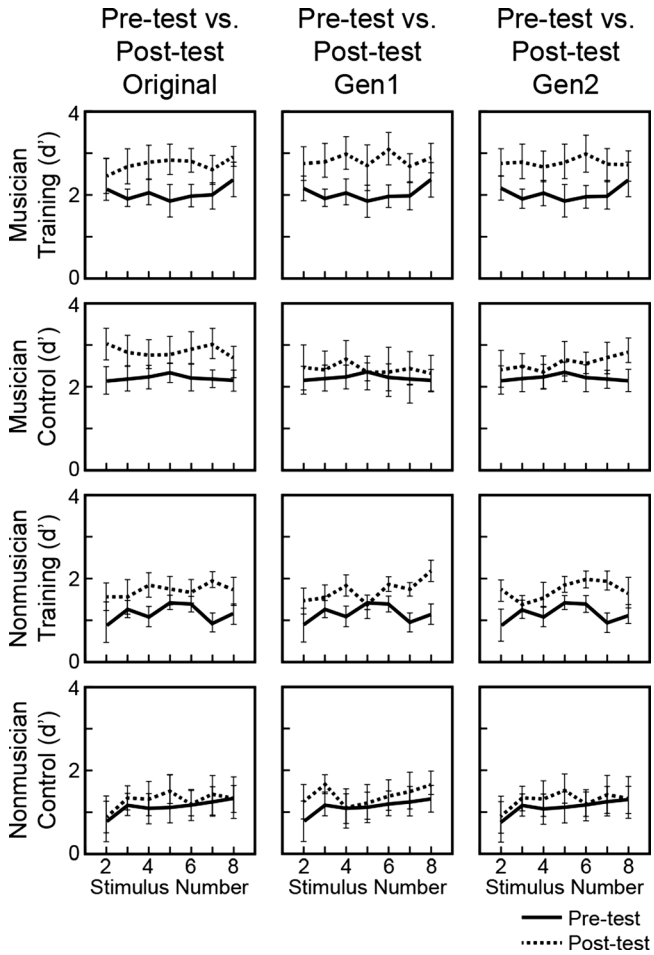

FIG. 7. Comparisons of within-pair sensitivities across the continuum between pre-training test (Experiment I) and three post-training tests (i.e., post-test with original stimuli, post-test with generalization stimuli 1, and post-test with generalization stimuli 2).

3-way mixed ANOVAs: Two musicianship (musicians, nonmusicians) by two training (trainees, controls) by two time (pre-training, post-training, repeated measure). For post-training tests with original stimuli and generalization 1 stimuli, only the main effects of time were significant, suggesting improvement in overall sensitivities among all groups [original stimulus: $F(1,36) = 44.94$, $p < 0.001$, $\eta_p^2 = 0.56$; generalization 1: $F(1,36) = 36.00$, $p < 0.001$, $\eta_p^2 = 0.29$]. For post-training test with generalization 2 stimuli, a training/time interaction was also observed in addition to main effect of time, suggesting a higher level of improvement specific to the training group [interaction: $F(1,36) = 5.10$, $p = 0.03$, $\eta_p^2 = 0.12$; main effect: $F(1,36) = 17.78$, $p < 0.001$, $\eta_p^2 = 0.33$, Fig. 8(A)].

b. Identification. Similarly, a separate and identical analysis was performed for each post-test stimulus to examine changes in slope and crossover measures related to training and music experience. All results are reported together here. For the slope measure, a three-way mixed ANOVA was conducted for each post-test: Two (musicianship: musicians vs nonmusicians) by 2 (training: Trainees vs controls) by 2 (time: Pre-training vs post-training, repeated measure). For post-training tests with original stimuli and generalization 1 stimuli, significant time effects were observed,

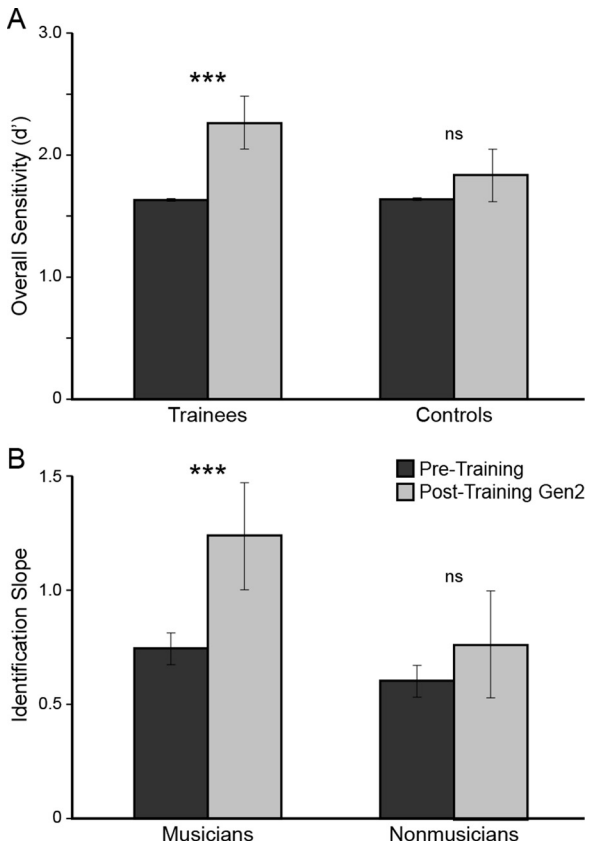

FIG. 8. A) Comparison of overall sensitivity (d') between pre-training test (Experiment I) and post-test with generalization stimuli 2 for four different groups (i.e., musician trainees, musician controls, nonmusician trainees, and nonmusician controls). The training by time interaction was observed. (B) Comparison of identification slope between pre-training test (Experiment I) and the post-training test with generalization stimuli 2 for the four different groups. An additional musicianship by time interaction was revealed.

suggesting increase of slope in all groups [original stimulus: $F(1,36) = 29.9$, $p < 0.001$, $\eta_p^2 = 0.45$; generalization 1: $F(1,36) = 20.1$, $p < 0.001$, $\eta_p^2 = 0.36$]. For post-training test with generalization 2 stimuli, a musicianship/time interaction was observed, $F(1,36) = 6.27$, $p = 0.017$, $\eta_p^2 = 0.15$, in addition to the main effect of time, suggesting a higher level of increase in slope specific to the musicians, $F(1,36) = 25.3$, $p < 0.001$, $\eta_p^2 = 0.41$. For the crossover measure, no significant effects were observed in any of the comparisons.

C. Discussion

Experiment II examined the effects of musical experience on the perceptual learning process of lexical tone categories. We randomly selected half of the English-speaking musicians and the English-speaking nonmusicians from Experiment I to complete a perceptual training procedure for the lexical tones categories of interest while the other half of the participants were controls. We then compared participants' perception pre-training to post-training among conditions (i.e., musician trainees, musician controls, nonmusician trainees, and nonmusician controls). We predicted that the perceptual training procedure would change the perception of the tone continuum in trainees. Differences in the change in perception between musicians and nonmusicians would provide evidence that musical experience influences learning lexical tone categories.

Discrimination data suggested that: (1) Training resulted in increased overall sensitivity across the continuum, even under the more difficult generalization condition, but not in the formation of robust categories; and (2) musicians and nonmusicians showed similar improvement. No within-pair sensitivity peak emerged on the tone continuum in any of the three post-training tests, suggesting that robust phonemic categories were absent after perceptual training. Analysis of changes in overall sensitivity between pre-training and post-training showed that participants in all conditions exhibited improvement in post-training tests with the original stimuli and generalization stimuli 1 (same talker, new syllable), suggesting a large practice effect by repeating the tasks. However, in the post-training test with generalization stimuli 2 (new talker, new syllable), only trainees (musicians and nonmusicians) showed an advantage in overall sensitivity improvement.

The conclusion that musicians and nonmusicians showed similar improvement was supported by discrimination data as well as identification data from the perceptual training. Over the 8-session perceptual training, both musician trainee and nonmusician trainee groups improved. However, no difference between groups was observed. Similarly, no group differences between musicians and nonmusicians were observed in the comparison between pre-training discrimination and post-training discrimination sensitivity.

Results from identification showed a slightly different picture. Participants in all conditions exhibited improvement in post-training tests with the original stimuli and generalization stimuli 1 (same talker, new syllable), similar to results from discrimination data. However, in the post-training test with the generalization stimuli 2 (new talker, new syllable), musicians (trainees and controls) showed improvement in identification slope, but not nonmusicians (trainees and controls).

We observed a significant increase in overall sensitivity and identification slope in all conditions. We speculate that the overall improvement observed in post-training tests is due largely to the orientation of attention to the cue of pitch contour. By tuning in to the correct cue, participants were able to devote more attentional resources and short-term memory to detecting and encoding the pitch component of the speech stimuli, resulting in better discrimination sensitivity and identification slope. While all participants were able to tune in to the pitch contours in the post-training tests with original stimuli and generalization stimuli 1, it was more difficult to stay oriented to the pitch contour cue with generalization stimuli 2. In this most difficult condition, in which the pitch contours were carried by a new vowel and a new speaker, perceptual training helped trainees stay oriented to the correct cue in the discrimination task, and musical experience helped musicians stay oriented in the identification task.

Other researchers have suggested that the ability to orient oneself to the critical cue is important in perceptual

learning. Goldstone (1998) suggested that there are four mechanisms contributing to perceptual learning, and the first one being "attentional weighting." In his description, "one way in which perception becomes adapted to tasks and environments is by increasing the attention paid to perceptual dimensions and features that are important, and/or by decreasing attention to irrelevant dimensions and features" (Goldstone, 1998, p. 588). In another recent study, Pederson and Guion-Anderson (2010) also provided evidence that orientation to the correct feature facilitates learning of speech sound contrasts.

Even though there is evidence in this study that some perceptual learning has taken place, none of the groups developed robust phonemic categories (i.e., "native-like" pattern of perception: Reduced sensitivity within category compared to sensitivity between categories). Why did training fail to result in a native-like pattern of perception? First of all, as presented in Experiment I, the training stimuli contained high variability in multiple dimensions: *F*0 (fundamental frequency) as well as other dimensions (voice quality, etc.). Even though *F*0 information has been demonstrated to be the most salient cue used by native speakers to categorize tones, other dimensions have influence as well, particularly for T3 (Wu, 2011). It is possible that the English-speaking learners relied on cues that are different from the native speakers such that in the post-testing stimuli, where only *F*0 dimension existed, categorization effects cannot be detected. Future research is warranted to: (1) Exert more control in matching of the dimensions of stimuli in training and testing, and (2) compare the results with learning from naturalistic multidimensional samples.

One other potential way to enhance the training paradigm is to provide additional distributional information of the pitch contour (lexical tone) categories in multiple modalities (i.e., vision). Although it has been shown that infants can take the distributional information from one modality from as short as 2 min of exposure (Maye *et al.*, 2002), such ability may be significantly reduced in adults. Although we provided distributional information in the auditory modality with high variability (multi-speaker and multi-tokens), it may require additional information (e.g., visual information) to facilitate phonemic category formation in adults. For example, training paradigms that involve multiple modalities show greater effects in training outcomes in non-native speech sound identification, music pitch discrimination, and detection of temporal violation (Hardison, 2003; Lappe *et al.*, 2008, 2011).

Taken together, the current study suggests that in adult learners, short-term perceptual training can result in perceptual learning regarding the attentional weighting of an important feature. However, we speculate that the formation of robust phonemic categories in adults might further require robust natural language input that involves multi-model information (e.g., visual information, motor information viewed in the vocal tract, etc.).

Finally, why did our results show limited differences between musicians and nonmusicians in the perceptual learning process? Previous research provided strong evidence that individuals with long-term musical experience exhibit robust advantages in encoding and detecting subtle differences in foreign lexical tones. Our results are consistent with the previous findings in that English-speaking musicians overall exhibited much higher sensitivities between lexical tone pairs. However, our results provided limited evidence that musicians exhibited advantages in the lexical tone category learning process. Similar patterns of improvement were observed in musicians and nonmusicians in the perceptual training tasks (2AFC and categorical discrimination) and in discrimination sensitivity on the tone continuum. Musicians exhibited an advantage only in the identification task. We speculate that the learning of phonemic categories might be a process that is largely independent from the processing of ordinary acoustic information. Learning phonemic categories involves higher-level cognitive processes, such as attentional networks, working memory, information integration, etc. Therefore, the effect of musical training may be limited to high acuity in information encoding and high levels of sensitivity to subtle acoustic differences, which may be beneficial only to some extent, for example, when the acoustic environment is less than optimal (e.g., noise-masked).

IV. GENERAL DISCUSSION

Previous research suggests that musical experience benefits individuals in processing and encoding foreign lexical tones. This study further examined the effects of musical experience on the perceptual learning process of lexical tone categories in two experiments. In Experiment I, we examined whether English-speaking musicians' perception of lexical tones are similar to English-speaking nonmusicians' when neither group had prior experience with tonal languages. Comparison of the perception of a tone continuum among English-speaking musicians, English-speaking nonmusicians, and Mandarin-speaking nonmusicians revealed that while musicians demonstrated the highest sensitivity to acoustic differences among stimuli, they exhibited a pattern of perception similar to the English-speaking nonmusicians and different from Mandarin-speaking nonmusicians. In Experiment II, we examined whether musical experience influences the perceptual learning of lexical tone categories. Our results suggest that while participants demonstrated some level of perceptual learning (attentional weighting), there was no strong evidence of improved learning specific to musicians. We speculate that the process of learning phonemic categories is largely independent of processing acoustic information. Future research is required to expand on this issue through several approaches: (1) To characterize the dominance/relative contribution of acoustic processing (bottom-up) versus phonemic categories (top-down) by examining musically-trained native Mandarin speakers; (2) to elucidate components that are necessary in the formation of robust phonemic categories, e.g., time of exposure, information from other modalities; (3) to optimize training paradigms that induce category learning in a targeted dimension (*F*0); (4) to investigate whether advantageous sensory processing benefits learning in less than optimal environments (e.g., noise-masked); and (5) to study whether musical

experience influences the learning process for foreign speech contrasts differently in children and adults.

ACKNOWLEDGMENTS

This study was supported by the NSF LIFE Science of Learning Center Program grant to the UW LIFE Center (P.K.K., PI: Grant No. SMA-0835854) and the UW Institute for Learning and Brain Sciences Ready Mind Project.

[1]http://lingua.mtsu.edu/chinese-computing/statistics/char/list.php?Which = mo

Alexander, J. A., Wong, P. C. M., and Bradlow, A. R. (**2005**). "Lexical tone perception in musicians and non-musicians," Paper presented at the *(Eurospeech) 9th European Conference on Speech Communication and Technology*, Lisbon, Portugal.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (**2008**). "Mixed-effects modeling with crossed random effects for subjects and items," J. Mem. Lang. **59**(4), 390–412.

Bates, D. M., Maechler, M., and Bolker, B. (**2012**). "lme4: Linear mixed-effects models using S4 classes," R package version 0.98-994.

Bidelman, G. M., Gandour, J. T., and Krishnan, A. (**2011**). "Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem," J. Cogn. Neurosci. **23**(2), 425–434.

Blicher, D. L., Diehl, R. L., and Cohen, L. B. (**1990**). "Effects of syllable duration on the perception of the Mandarin tone-2 tone-3 distinction—evidence of auditory enhancement," J. Phonetics **18**(1), 37–49.

Boersma, P., and Weenink, D. (**2009**). "Praat: Doing phonetics by computer," (Version 5.1.05) [Computer program] (Last viewed August 2013).

Braida, L. D., Lim, J. S., Berliner, J. E., Durlach, N. I., Rabinowitz, W. M., and Purks, S. R. (**1984**). "Intensity perception. 13. Perceptual anchor model of context-coding," J. Acoust. Soc. Am. **76**(3), 722–731.

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), 1–9.

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 (L)," J. Acoust. Soc. Am. **119**(2), 719–722.

Durlach, N. I., and Braida, L. D. (**1969**). "Intensity perception. 1. Preliminary theory of intensity resolution," J. Acoust. Soc. Am. **46**(2P2), 372–383.

Francis, A. L., Ciocca, V., and Ng, B. K. C. (**2003**). "On the (non)categorical perception of lexical tones," Percept. Psychophys. **65**(7), 1029–1044.

Goldstone, R. L. (**1998**). "Perceptual learning," Annu. Rev. Psychol. **49**, 585–612.

Halle, 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**(3), 395–421.

Hardison, D. M. (**2003**). "Acquisition of second-language speech: Effects of visual cues, context, and talker variability," Appl. Psycholinguist. **24**(04), 495–522.

Harnad, S. R. (**1987**). *Categorical Perception: The Groundwork of Cognition* (Cambridge University Press, Cambridge, New York).

Hyman, L. M. (**2001**). "Tone systems," in *Language Topology and Language Universals: An International Handbook*, edited by M. Haspelmath (W. de Gruyter, Berlin), Vol. 2.

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**(1), B47–B57.

Kiriloff, C. (**1969**). "On auditory perception of tones in Mandarin," Phonetica **20**(2–4), 63–67.

Kraus, N., and Chandrasekaran, B. (**2010**). "Music training for the development of auditory skills," Nat. Rev. Neurosci. **11**(8), 599–605.

Kuhl, P. K., and Iverson, P. (**1995**). "Linguistic experience and the 'perceptual Magnet effect,'" in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York Press, Baltimore, MD).

Lappe, C., Herholz, S. C., Trainor, L. J., and Pantev, C. (**2008**). "Cortical plasticity induced by short-term unimodal and multimodal musical training," J. Neurosci. **28**(39), 9632–9639.

Lappe, C., Trainor, L. J., Herholz, S. C., and Pantev, C. (**2011**). "Cortical plasticity induced by short-term multimodal musical rhythm training," PLoS One **6**(6) e21493.

Lee, C. Y., and Hung, T. H. (**2008**). "Identification of Mandarin tones by English-speaking musicians and nonmusicians," J. Acoust. Soc. Am. **124**(5), 3235–3248.

Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (**1957**). "The discrimination of speech sounds within and across phoneme boundaries," J. Exp. Psychol. **54**(5), 358–368.

Logan, J. S., Lively, S. E., and Pisoni, D. B. (**1991**). "Training Japanese listeners to identify English /r/ and /l/: A 1st report," J. Acoust. Soc. Am. **89**(2), 874–886.

Macmillan, N. A., and Creelman, C. D. (**2008**). *Detection Theory: A User's Guide* (Psychology Press, New York).

Macmillan, N. A., Goldberg, R. F., and Braida, L. D. (**1988**). "Resolution for speech sounds: Basic sensitivity and context memory on vowel and consonant continua," J. Acoust. Soc. Am. **84**(4), 1262–1280.

Marques, C., Moreno, S., Castro, S. L., and Besson, M. (**2007**). "Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence," J. Cogn. Neurosci. **19**(9), 1453–1463.

Maye, J., Werker, J. F., and Gerken, L. (**2002**). "Infant sensitivity to distributional information can affect phonetic discrimination," Cognition **82**(3), B101–B111.

Patel, A. D. (**2011**). "Why would musical training benefit the neural encoding of speech? The OPERA hypothesis," Frontiers Psychol. **2**, 142.

Patel, A. D., and Iversen, J. R. (**2007**). "The linguistic benefits of musical abilities," Trends Cogn. Sci. **11**(9), 369–372.

Pederson, E., and Guion-Anderson, S. (**2010**). "Orienting attention during phonetic training facilitates learning," J. Acoust. Soc. Am. **127**(2), EL54–EL59.

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**(4), 616–624.

R Development Core Team (**2012**). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org.

Schon, D., Magne, C., and Besson, M. (**2004**). "The music of speech: Music training facilitates pitch processing in both music and language," Psychophysiology **41**(3), 341–349.

Shen, X. N. S., and Lin, M. C. (**1991**). "A perceptual study of Mandarin Tone 2 and Tone 3," Lang. Speech **34**, 145–156.

Wang, Y., Spence, M. M., Jongman, A., and Sereno, J. A. (**1999**). "Training American listeners to perceive Mandarin tones," J. Acoust. Soc. Am. **106**(6), 3649–3658.

Wayland, R. P., and Li, B. (**2008**). "Effects of two training procedures in cross-language perception of tones," J. Phonetics **36**(2), 250–267.

Wing, H. D. (**1966**). *Manual for Standardized Tests of Musical Intelligence* (City of Sheffield Training College, Sheffield, England).

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**(4), 420–422.

Wu, J. (**2011**). "A study on the perceptual space of Putonghua Tone 2 and Tone 3," Master thesis, Peking University. Retrieved from http://cdmd.cnki.com.cn/Article/CDMD-10001-1011132278.htm (Last viewed April 3, 2014).

Xu, Y. S., 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**(2), 1063–1074.

Yang, R. X. (**2011**). "The phonation factor in the categorical perception of Mandarin tones," Paper presented at the *ICPhS XVII*, Hong Kong.

Zatorre, R. J., and Gandour, J. T. (**2008**). "Neural specializations for speech and pitch: Moving beyond the dichotomies," Philos. Trans. R. Soc., B **363**(1493), 1087–1104.

Zhao, T., Wright, R., and Kuhl, P. (**2012**). "Modeling Mandarin Tone 2 and Tone 3 from natural productions with variability," J. Acoust. Soc. Am. **131**(4), 3346.