The effects of musicality and language background on cue integration in pitch perception

Aletheia Cui and Jianjing Kuang[a)]

Department of Linguistics, University of Pennsylvania, 3401-C Walnut Street, Suite 300, Philadelphia, Pennsylvania 19104, USA

(Received 7 June 2019; revised 23 October 2019; accepted 28 October 2019; published online 3 December 2019)

Pitch perception involves the processing of multidimensional acoustic cues, and listeners can exhibit different cue integration strategies in interpreting pitch. This study aims to examine whether musicality and language experience have effects on listeners' pitch perception strategies. Both Mandarin and English listeners were recruited to participate in two experiments: (1) a pitch classification experiment that tested their relative reliance on f0 and spectral cues, and (2) the Montreal Battery of Evaluation of Musical Abilities that objectively quantified their musical aptitude as continuous musicality scores. Overall, the results show a strong musicality effect: Listeners with higher musicality scores relied more on f0 in pitch perception, while listeners with lower musicality scores were more likely to attend to spectral cues. However, there were no effects of language experience on musicality scores or cue integration strategies in pitch perception. These results suggest that less musical or even amusic subjects may not suffer impairment in linguistic pitch processing due to the multidimensional nature of pitch cues. © *2019 Acoustical Society of America*.

https://doi.org/10.1121/1.5134442

[ZZ] Pages: 4086–4096

I. INTRODUCTION

Pitch is not only an essential element of music but also conveys crucial contrasts in speech (e.g., tone and intonation). The relationship between pitch processing in speech and music is complex. While the auditory processing of music and speech tends to lateralize to the right and left hemispheres, respectively (Bever, 1975; Zatorre *et al.*, 2002), music and speech also share neural processing mechanisms (Bidelman *et al.*, 2011; Koelsch *et al.*, 2002; Patel *et al.*, 1998; Patel *et al.*, 2008; Peretz *et al.*, 2015). It is possible that there is transfer between linguistic and musical pitch processing. In particular, musicality, the cognitive ability to process music, has been found to significantly affect human listeners' performance on linguistic pitch perception. However, the effect of language experience on pitch perception is less clear. It has been of particular interest whether tone language speakers may have an advantage in pitch perception, and the results so far are inconclusive. Moreover, an additional layer of complexity arises since listeners integrate multiple acoustic cues (especially f0 and spectral cues) in pitch perception (e.g., Kuang and Liberman, 2018). The question remains as to what extent musicality and tone language experience influence listeners' cue integration strategies when processing pitch.

A. Musicality and language processing

Musicality refers to an individual's aptitude for processing various aspects of music. Part of musicality can be attributed to innate predisposition. On one end of the scale, there are individuals born with congenital amusia, who have difficulty discriminating pitch differences (Peretz *et al.*, 2002). On the other end, some children are able to acquire absolute pitch with early music training (Chin, 2003). Family-aggregation studies suggest that there is a genetic basis for amusia and the predisposition to acquire absolute pitch (Baharloo *et al.*, 2000; Gregersen *et al.*, 2001; Peretz *et al.*, 2007). In addition to genetic predisposition, musical training can lead to significantly better auditory processing and enhanced musical abilities (Fujioka *et al.*, 2006; Kraus and Chandrasekaran, 2010; Peretz *et al.*, 2013). Musicality, therefore, reflects the cognitive capability for music shaped by both genetic background and learning.

There is extensive evidence for a significant relationship between good musicality and enhanced linguistic processing. For example, better musicality is associated with better pitch processing in speech (e.g., Moreno *et al.*, 2008; Schön *et al.*, 2004), reading skills (e.g., Moreno *et al.*, 2008), speech segmentation (e.g., François *et al.*, 2012), and second language learning (e.g., Posedel *et al.*, 2012). Among non-tone language speakers, musicians outperform non-musicians at discriminating and learning lexical tone contrasts (Alexander *et al.*, 2005; Lee and Hung, 2008; Marie *et al.*, 2011; Wong and Perrachione, 2007). Moreover, children and adults with music training are better at identifying pitch contour distinctions in both music and speech (Magne *et al.*, 2006; Schön *et al.*, 2004).

Since musicality is beneficial for language learning and processing, one might expect that people with poor musicality and especially people with congenital amusia would face major disadvantages in language processing, particularly for tone and intonation. However, it remains an open question how much language processing can be affected by amusia. For example, although amusic subjects are impaired in

---

[a)] Electronic mail: kuangj@ling.upenn.edu processing the pitch of musical melodies (Kalmus and Fry, 1980), amusic tone language speakers typically experience little trouble in tone perception (Liu *et al.*, 2016; Nan *et al.*, 2010). Moreover, only some of the amusic tone language speakers also have tone agnosia, or impairment in tone perception, and even these speakers generally exhibit normal tone production (Liu *et al.*, 2016; Nan *et al.*, 2010). An interesting question thus arises: How are these speakers able to acquire normal tone production if they struggle to perceive pitch distinctions?

This discrepancy between perception and production led to the proposal that the perception and production of pitch have different neural pathways (Hutchins and Moreno, 2013; Liu *et al.*, 2016; Nan *et al.*, 2010). Complete disassociation between pitch perception and production during acquisition is unlikely, since it would be difficult for amusic learners to develop reliable production without appropriate auditory feedback of their pitch production. In addition, Nan *et al.* (2010) considers the possibility that amusic individuals might use non-pitch based cues to guide their production of pitch. This is plausible because speech cues are often highly redundant. However, it may not be necessary to incorporate non-pitch cues, as pitch perception and production are not solely determined by the fundamental frequency (f0), and they are in fact mapped onto multiple acoustic cues (cf. Sec. I C). Therefore, in this study, we investigate a third possibility: People who are less sensitive to f0 differences can use acoustic cues that co-vary with f0 to perceive tonal pitch distinctions—that is, musicality may have effects on cue sensitivity in pitch perception. Since musicality within a population forms a continuous spectrum, we further hypothesize that this cognitive variation in pitch cue integration is not limited to people with neurogenetic disorders but can also be observed among the normal hearing population.

B. Language experience and pitch processing

While previous findings show consistent effects of musicality on linguistic processing, the influence of language background on musical abilities is more complicated. Since tone language speakers are able to reliably produce consistent pitch differences for lexical contrasts, their experience with linguistic pitch may have an effect on musical pitch processing. Although tone language speakers generally outperform non-tone language speakers in lexical tone perception tasks (e.g., Bent *et al.*, 2006; Francis *et al.*, 2008; So and Best, 2010; Wayland and Guion, 2004), it is controversial whether this effect transfers to general pitch processing.

On the one hand, there is evidence that tone language speakers have an advantage in general pitch processing. For example, tone language speakers are more likely to exhibit absolute pitch (Deutsch *et al.*, 2009; Deutsch *et al.*, 2006), have more categorical perception of non-speech pitch contours (Peng *et al.*, 2010; Xu *et al.*, 2006), and outperform non-tone languages speakers on various musicality tests (Chen *et al.*, 2016; Pfordresher and Brown, 2009; Stevens *et al.*, 2013). On the other hand, some studies have suggested that speaking a tone language does not necessarily lead to enhanced pitch perception. For example, tone language speakers have similar rates of amusia (Kalmus and Fry, 1980; Nan *et al.*, 2010) and similar patterns of musical impairment as non-tone language amusic subjects (Peretz *et al.*, 2013). Moreover, recent studies found that the benefit of language experience appears to only transfer to linguistic tasks but not to general pitch perception (Bent *et al.*, 2006; DiCanio, 2012; Ngo *et al.*, 2016; Zheng and Samuel, 2018). These results have led some authors to argue for a "specific transfer" view for the relationship between pitch perception in language and music: Language experience only enhances performance in speech pitch tasks but does not transfer to general musical abilities (Ngo *et al.*, 2016; Zheng and Samuel, 2018).

Overall, there appears to be some language effect on pitch perception, but this effect is rather complex, and perhaps domain specific (e.g., speech vs non-speech) and/or task specific (e.g., identification vs discrimination). In addition to whether language experience has effects on the sensitivity to pitch differences, another important issue is to what extent language experience can affect the use of acoustic cues in pitch perception. Language experience is known to significantly modulate the dimensionality and relative cue weighting of the perception of linguistic categories (e.g., Escudero *et al.*, 2009; Holt and Lotto, 2006; Iverson *et al.*, 2003; Lipski *et al.*, 2012). Specific to linguistic tones, studies have focused on two main f0 cues that are involved in the tonal contrasts: f0 height (e.g., high vs low) and the direction of f0 contours (e.g., rising vs falling). The relative importance of these dimensions are shaped by the speakers' language experience (Gandour, 1983; Gandour and Harshman, 1978; Xu *et al.*, 2006): While f0 height is important across languages, f0 direction is relatively more important to speakers of tone languages, especially those of East and Southeast Asia. However, it remains an open question to what extent language experience affects cue weighting in general pitch perception (Chen *et al.*, 2018).

C. Acoustic cues for pitch perception

Most studies on music and linguistic pitch processing focus primarily on f0 (e.g., Alexander *et al.*, 2005; Bidelman *et al.*, 2013; Schön *et al.*, 2004; Wong and Perrachione, 2007). However, the interpretation of pitch can and does often involve multiple cues, and listeners vary in their reliance on these cues. For example, even when f0 perception is considered by itself, at least two strategies can be used. By definition, f0 is the lowest frequency component of a periodic waveform, and it can be found by identifying the lowest component in the Fourier-transformed spectrum. In addition, it can be perceived through resolving the frequency intervals between the harmonics. Therefore, in missing fundamental experiments, where the first harmonic is removed from the spectrum and only the higher harmonics are available, listeners exhibit individual variation in pitch perception (e.g., Ladd *et al.*, 2013; Seither-Preisler *et al.*, 2007): Some listeners perceived the missing f0 by resolving the intervals of present harmonics, while other listeners perceived the lowest present harmonic as the f0.

Moreover, pitch perception also integrates a number of non-f0 cues, such as temporal envelope and temporal fine structure (e.g., Kong and Zeng, 2006), spectral shape (e.g., Allen and Oxenham, 2014; Kuang and Liberman, 2018), periodicity/roughness (Kuang and Liberman, 2016), and intensity (Neuhoff and McBeath, 1996). The present study focuses on spectral cues. In the perception of non-speech pitch, there is significant interaction between f0 and different spectral shapes (i.e., timbre), such as the natural differences between musical instruments or the synthetically manipulated center frequencies in complex tones (e.g., Allen and Oxenham, 2014; Krumhansl and Iverson, 1992; Marozeau and de Cheveigné, 2007; Melara and Marks, 1990, to name a few). A common finding from this body of research is that in speeded pitch classification tasks, listeners' pitch perception is more accurate and faster when the spectral/timbre cues are consistent with the corresponding f0 dimension.

Additionally, spectral slope, defined as the relative amplitude difference between the low frequency components and the high frequency components in the spectrum, can be a linguistically meaningful acoustic cue. It is one of the most important acoustic correlates of voice quality, broadly defined as the configuration and tension settings of vocal fold vibration during voicing (see Gobl and Ní Chasaide, 2013, for a detailed review). Due to the physiological co-variation between f0 and voice quality (Titze, 1988; Zhang, 2016), spectral slope systematically co-varies with f0 in pitch production (Kuang, 2017). Specifically, when producing a high pitch, the vocal folds are naturally more constricted and have greater longitudinal tension, and the voice quality is thus much tenser; however, when producing a pitch target in the speaker's comfortable range, the vocal folds are usually less constricted, resulting in a breathier sound. Note that the terms we use here are relative since voice quality varies along with the f0 scale in a continuous manner. Acoustically, tenser voice has a flatter spectral slope and breathier voice has a steeper spectral slope (cf. Figure 1).

The co-variation between voice quality and f0 is also observed in pitch perception with speech stimuli. Previous studies have shown that the manipulation of spectral slope can cause significant shift in pitch perception, with flatter slope (i.e., tenser voice) perceived as being higher in pitch than steeper slope at the same f0 (Kuang and Liberman, 2015, 2018). Notably, Kuang and Liberman (2015) found that listeners differed in their reliance on spectral slope and f0 cues in pitch perception. Specifically, non-musicians relied more on spectral slope cues while musicians relied more on f0. This pattern is similar to the variation found in the missing fundamental experiments (Seither-Preisler *et al.*, 2007), in which subjects with higher musical competence categorized stimuli based on the missing f0 rather than the frequencies of the overtone spectra. Therefore, we hypothesize that musicality can be predictive of listeners' strategies in pitch perception: Listeners with better musicality are more likely to rely on f0, while listeners with worse musicality are more likely to pay attention to cues that co-vary with f0.

D. The current study

This study aims to investigate whether musicality has effects on cue integration (specifically f0 and spectral slope) in pitch perception, and whether these effects are modulated by tone language background. To address these questions, we ran a pitch classification experiment and a musicality test on non-tone language (English) speakers and tone language (Mandarin) speakers. The pitch classification experiment is the same as Kuang and Liberman (2018), in which listeners classified the relative pitch of an f0 continuum under four different spectral slope conditions. An earlier study with a similar experimental protocol found evidence for an effect of musicianship on pitch perception based the subjects' self-reported musicianship status (Kuang and Liberman, 2015). Unlike this earlier study, the current study objectively quantifies the subjects' musical aptitude using a standard musicality test. Although music training is associated with improved performance on music-related tasks (Bidelman *et al.*, 2013; Peretz *et al.*, 2013), there is individual variation in innate musical talent. Among the population without music training, there are people who have a natural aptitude for music, termed "musical sleepers" by Law and Zentner (2012).

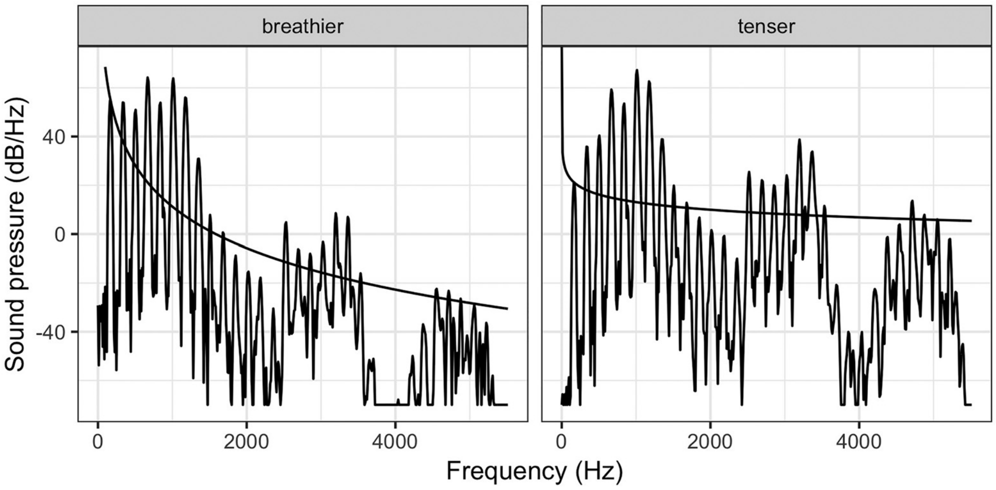

FIG. 1. Left: the spectral slope of the natural phonation of the speaker. Right: the spectral slope is flattened to create tenser-sounding phonation.

Therefore, to better understand the relationship between music and speech, it is preferable to obtain objective measures of musicality rather than relying on self-reported music training.

By comparing the results from the pitch perception experiment and the musicality experiment by language group, it is possible to identify any language effect on musicality and pitch processing. In addition, an analysis on the individual level is carried out on the relationship between musicality and pitch perception strategies in order to explore the role of musical aptitude in linguistic pitch processing. Overall, this study aims to provide a better understanding of the relationship between native language, musicality, and the use of f0 and spectral cues in pitch perception.

II. METHODS

A. Participants

Seventy-one native English speakers (mean age 19.82; sd 1.66) and 44 native Mandarin speakers (grew up in China; mean age 24.95; sd 6.92) were recruited to participate in two perception experiments. The experimental protocol was approved by the University of Pennsylvania Institutional Review Board. All participants reported normal hearing. The subjects first completed a pitch perception experiment and then a musicality experiment. The pitch perception experiment was conducted first to avoid any potential training effect from completing the musicality experiment.

B. Experiment 1: Pitch perception

The pitch perception experiment, adapted from the procedures in Kuang and Liberman (2018), tests each subject's reliance on f0 and spectral slope cues in a relative pitch judgment task. The stimuli are sets of utterances with two f0 peaks differing in four spectral conditions, and they are resynthesized from a natural production of a "ma-MA-ma" utterance of a male speaker. The prosody of the utterance is the same as "banana," with a natural Low-High-Low pitch contour. The f0 starts low and rises through the first syllable, peaks on the middle syllable, and falls on the third syllable. The original phonation of the speaker constitutes the "breathier" voice quality in this experiment. A "tenser" version of the "ma-MA-ma" utterance was created using TANDEM-STRAIGHT (Kawahara *et al.*, 2008) so that the spectral slope of the Fourier spectrum was 6 dB/octave less steep than the breathier version. As illustrated in Fig. 1, the tenser version of the spectrum has an overall flatter spectral slope. Each stimulus consists of two continuous "ma-MA-ma" utterances in each of the four possible voice quality combinations (Table I). For example, for the breathier-tenser (BT) condition, the listener would hear a breathier "ma-MA-ma" followed immediately by a tenser "ma-MA-ma" with no pause in between.

TABLE I. Summary of the four spectral slope conditions.

Peak 1	Peak 2	Intended voice quality combination
Tilted	Tilted	Breathier + Breathier (BB)
Flat	Flat	Tenser + Tenser (TT)
Tilted	Flat	Breathier + Tenser (BT)
Flat	Tilted	Tenser + Breathier (TB)

In addition to spectral slope manipulations, the f0 contour is also modified (Fig. 2). The lowest f0 is the same for both "ma-MA-ma" phrases (120 Hz). While the maximum value of the first peak is kept constant at 169.34 Hz, the second peak is an 11-step continuum varying from 153.06 to 187.36 Hz (0.35 semitone/step). At step 6, the f0 contour of the second phrase is the same as the f0 contour on the first phrase. After the spectral and f0 manipulations, the amplitude of the two "ma-MA-ma" phrases is normalized by scaling the maximum amplitude of each phrase to 1. In total, there are 44 distinct stimuli (4 spectral slope conditions × 11 f0 steps).

The listeners participated in this experiment in a soundproof booth. The stimuli were played through Sennheiser HD 280 Pro headphones. The subjects were instructed to think of each "ma-MA-ma" utterance as a word, and upon hearing a stimulus with two "ma-MA-ma" words, they were asked to do a forced-choice classification to indicate which of the two "words" sounded higher in pitch. Listeners were also instructed to refer to their linguistic knowledge of pitch, such as intonation (e.g., the question intonation in "Sure?" vs the declarative intonation in "Sure." for English speakers) or tone (e.g., high tone vs low tone for Mandarin speakers) to judge pitch height. The presentation of the stimuli was randomized, and each stimulus was presented five times.

The goal of this experiment is to test to what extent spectral conditions would cause subjects to shift their judgment of relative pitch. The term "shift" here is used to refer to any difference in pitch judgment along the f0 continuum as the result of the spectral slope conditions. For example, if subjects rely purely on f0 to identify relative pitch, there should not be any differences in their judgment patterns across the four spectral slope conditions. They should consistently choose the first peak as higher for f0 steps 1–5 across all spectral conditions, and second peak as higher for f0 steps 7–11 across all spectral conditions. However, if listeners are affected by spectral slope cues, shift should occur in their perception of pitch. They might choose the second peak as higher at lower f0 steps for the breathier-tenser condition, since tenseness is associated with higher pitch. Similarly, for the tenser-breathier condition, they might start choosing the second peak as higher at higher f0 steps. These

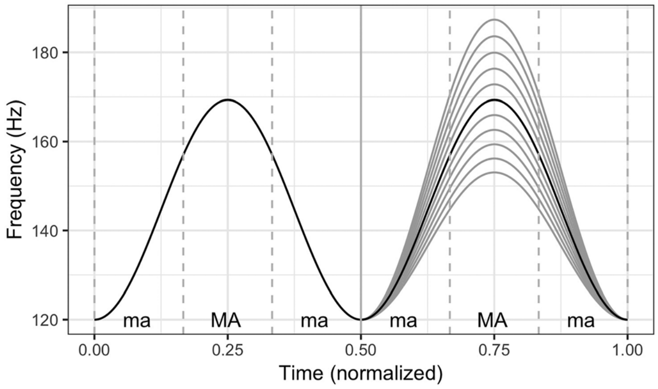

FIG. 2. The f0 contours for the stimuli of the two "ma-MA-ma" phrases. The second peak is a continuum of 11 f0 steps.

different predicted patterns are illustrated in Fig. 3, which shows the rates at which subjects are predicted to select the second peak as higher for the different spectral conditions when they do not exhibit shift (left) and when they do exhibit shift (right).

To quantitatively examine cue integration in pitch perception on the individual level, for each subject, a perceptual shift score and a categoricity score are calculated based on second peak higher response rates. The perceptual shift score is meant to measure how much a subject's responses are influenced by the spectral slope conditions. The more shifted subjects should exhibit greater distances between their response curves for the different spectral conditions. Categoricity measures how sharp the perceptual boundary is between two classification categories (first peak higher vs second peak higher). There are commonly used methods for estimating boundary shift and categoricity for binary categorization data. For example, appropriate classification functions, such as logistic functions, can be used to fit such categorization data. From the raw data or fitted functions, boundary shift can be measured as comparisons between the points where response rates cross over 50% for the different experimental conditions (e.g., Best *et al.*, 1981; Kuhl and Miller, 1978), and categoricity can be estimated as some measure of boundary width or the slope of the fitted function (e.g., Kuhl and Miller, 1978). The steeper the slope, the more categorical the responses are. In this study, mathematically simpler methods of estimating shift and categoricity are adopted because of the range of individual variation found during the preliminary analysis of the results. For example, while some subjects exhibit categorical perception along the 11 f0 steps that can be modeled with a logistic function, results from other subjects are more linear. Some subjects chose the second peak as higher for BT and first peak higher for tenser-breathier (TB) for almost all the f0 steps. For these subjects, the logistic function is a poor fit.

Specifically, in this study, perceptual shift between the spectral slope conditions is quantified for each subject as the mean of their BT second peak higher response rates minus the mean of their TB second peak higher response rates. Subjects who are more affected by the spectral slope conditions should have greater differences between their responses in the BT conditions and their responses in the TB condition. Categoricity of the responses is scored for each subject as the mean second peak higher response rates for f0 steps 7–11 minus the mean second peak higher response rates for f0 steps 1–5. The means are taken across all four spectral slope conditions. Subjects who exhibit more categorical responses should have higher scores, while subjects with more flat response curves should have lower scores.

C. Experiment 2: Musicality

Since music training is associated with better musicality, each participant first answered a brief survey about whether they had formal music training and for how long. To quantitatively evaluate musicality, the abbreviated Montreal Battery of Evaluation of Musical Abilities (MBEMA) was used (Peretz *et al.*, 2013). The MBEMA stimuli are short musical phrases played with different instruments. Each phrase is either played by itself or paired with another musical phrase depending on the task.

There are three tasks, evaluating the subject's ability to identify differences in melody and rhythm as well as their memory of the musical phrases. Each task consists of 20 stimuli, and there are 60 stimuli in total. The melody test plays two consecutive melodies and asks the subject to identify whether the melodies are the same or different. The rhythm test also plays two consecutive musical phrases, which may or may not have the same rhythm. Last, the memory test plays only one melody, and it asks the listener to indicate whether they have heard this melody in the previous tasks. For each subject, the score for each test is calculated as the percentage of correct answers for that test. An overall musicality score is calculated as the percentage of correct answers for all three tests.

III. RESULTS

A. Pitch perception

Figure 4 illustrates the pitch perception results by English and Mandarin speakers. The plot shows the rates at

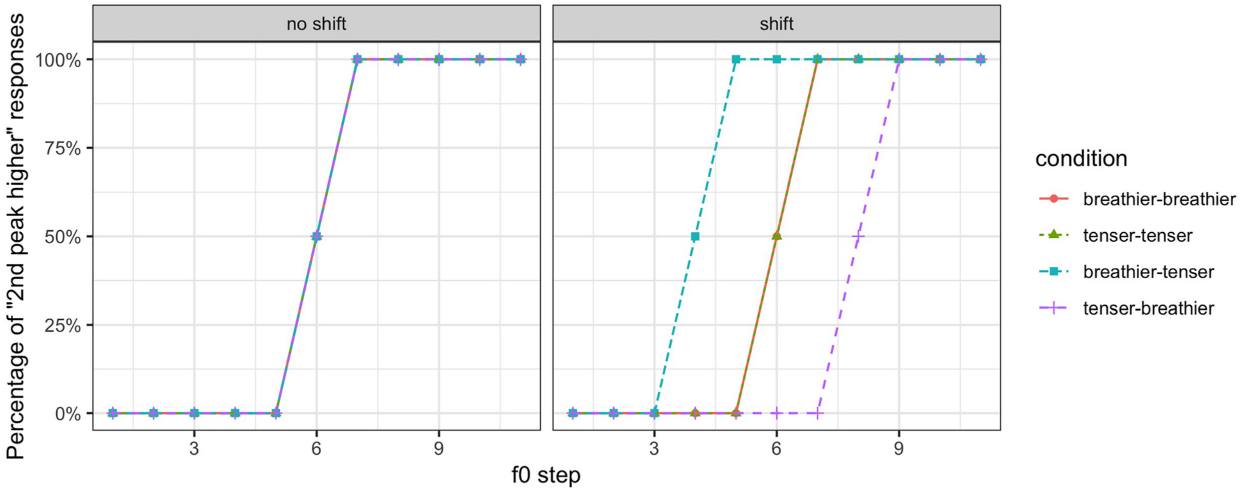

FIG. 3. (Color online) Left: Expected response pattern from listeners who are not affected by spectral cues in pitch judgment. Right: Expected response pattern from listeners who are affected by spectral cues in pitch judgment.

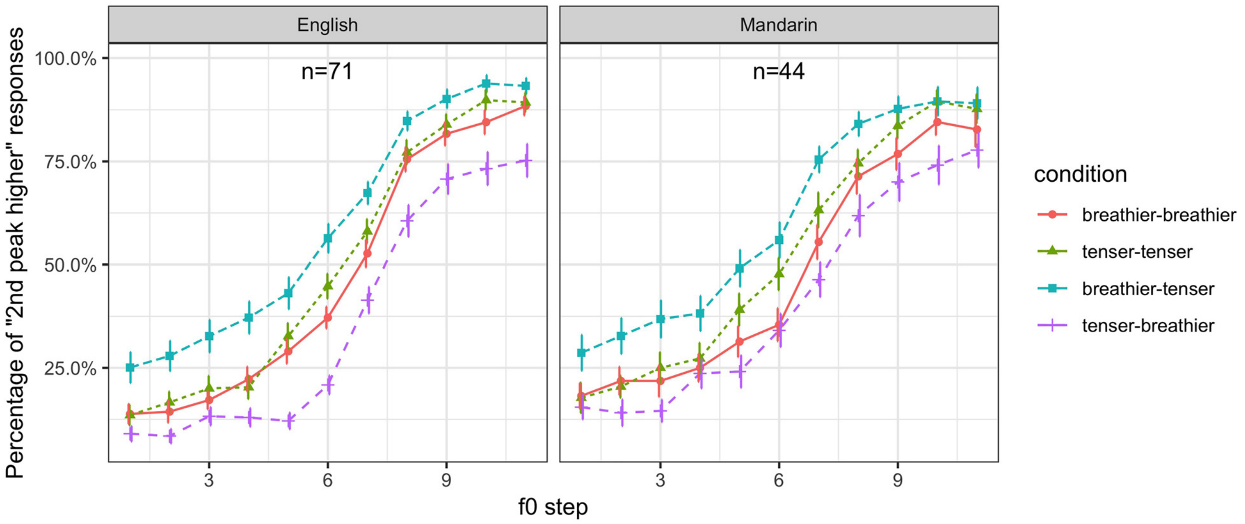

FIG. 4. (Color online) Second peak higher response rates by English and Mandarin speakers. X-axis = f0 steps, *y*-axis = proportion of responses where subjects chose peak 2 to be higher.

which subjects chose the second peak to be higher for the four spectral slope conditions. Tenser-tenser (TT) and breathier-breathier (BB) have the same spectral slope for both peaks and can serve as the reference conditions to observe whether TB and BT, which have different spectral slopes on the two peaks, have effects on pitch judgment. As can be seen from the figure, both language groups show similar shifts for the TB and BT spectral slope conditions. When the second peak has a tenser voice compared to the first (BT condition), listeners select the second peak as higher at lower f0 steps, showing that their pitch judgment is affected by the boosted frequencies in the tense peak. Likewise, when the second is breathier (TB condition), listeners select the second peak as higher in pitch at lower rates than either BB or TT.

For this experiment, the main effects of spectral slope, f0, and native language on second peak higher responses are evaluated using an MCMC generalized linear mixed-effects model in R (Hadfield, 2010). Spectral slope conditions (four factor levels), f0 steps (eleven numeric steps), and native language (two factor levels) are included as fixed effects, and subjects are the random intercepts. Pairwise comparisons between every two spectral slope conditions are achieved by changing the reference level. The main effects of spectral slope conditions are summarized in Table II. As expected, across all models, f0 is a significant predictor for subject responses (posterior mean = 0.36, 95% posterior density intervals = [0.33, 0.39], $p < 0.001$), while native language is not significant (posterior mean = 0.13, 95% posterior density intervals = [−0.01, 0.27], $p = 0.074$).

Overall, these results suggest that both Mandarin and English listeners integrate f0 and spectral slope cues in relative pitch judgment and replicate the pattern from Kuang and Liberman (2018). Quantitative analysis of perceptual shift and language differences is presented in Sec. III C.

B. Musicality and musicianship

All subjects scored relatively high on the musicality test. The combined mean for English and Mandarin speakers is 88.33%, with a standard deviation of 6.99. When separated into language groups, both English and Mandarin speakers have similar group means and standard deviations across all three tests and the overall scores (Table III). A Welch two sample t-test shows that the group means of the total scores are not significantly different between the two language groups ($t = 0.275$, $df = 93.156$, $p = 0.784$).

Figure 5 shows the distribution of musicality scores across all subjects, as well as the proportions of self-identified musicians vs non-musicians for each score. Subjects with music training ($N = 66$, mean = 2.70, sd = 0.17) have on average higher musicality scores than subjects without music training ($N = 48$, mean = 2.58, sd = 0.23). A Welch two sample t-test shows that the group means are significantly different between the subjects with musical training and the subjects without musical training ($t = 3.15$, $df = 83.12$, p-value = 0.002).

Out of the English group, 47 subjects have musical training and 24 do not. For the Mandarin group, only 20 subjects have musical training, and 24 do not. An analysis of variance (ANOVA) model is run to analyze the effects of native language, musicianship, and their interaction on the musicality test outcome. The main effect of music training is significant [$F(1,110) = 10.894$, MSE = 0.049, $p = 0.001$].

TABLE II. Main effects of spectral slope for every pair of conditions for all the speakers. Means of regression coefficients are followed by the 95% highest posterior density intervals in square brackets and the associated p-values.

	BB	TT	BT
TT	0.16 [0.07, 0.26], $p < 0.001$		
BT	0.59 [0.50, 0.69], $p < 0.001$	0.45 [0.36, 0.54], $p < 0.001$	
TB	−0.43 [−0.53, −0.35], $p < 0.001$	−0.62 [−0.72, −0.52], $p < 0.001$	−0.85 [−0.93, −0.78], $p < 0.001$

TABLE III. Summary statistics for the musicality test for the English and Mandarin groups, including results for the melody, rhythm, and memory subtests as well as the total scores.

	Melody		Rhythm		Memory		Total	
Language	mean	sd	mean	sd	mean	sd	mean	sd
English	88.14%	10.63	89.21%	8.99	88.07%	8.56	88.48%	7.09
Mandarin	86.81%	9.16	87.95%	8.85	89.54%	8.06	88.11%	6.92

Native language is not significant [F(1,110) = 0.081, MSE = 0.0004, n.s.], and the interaction of these two factors is also not significant [F(1,110) = 1.283, MSE = 0.006, n.s.].

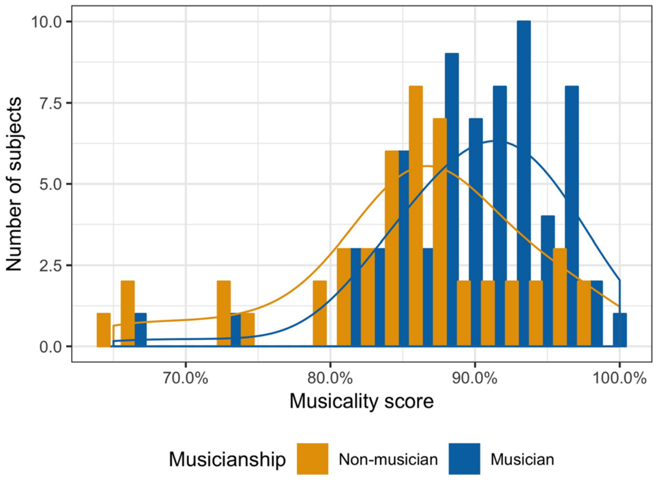

FIG. 5. (Color online) The distribution of musicality scores for musicians and non-musicians.

C. The relationship between musicality and pitch perception

To visualize the relationship between musicality and pitch perception strategies, subjects are separated into five quantiles by their musicality scores. Because many subjects share the same scores (Fig. 5), the cut off scores (85.00%, 88.33%, 90.00%, and 93.33%) for the quantiles do not divide the subjects into equal groups. There are 22, 23, 16, 19, and 34 subjects in each quantile from low to high. Figure 6 plots the pitch perception results for each group of subjects. Group 1 corresponds to subjects whose scores are below the 20th percentile, group 2 corresponds to subjects with scores between the 20th percentile and the 40th percentile, and so on. There are two major differences between the groups. First, subjects with low musicality scores show greater shift as the result of spectral slope conditions. For instance, subjects in group 1 are the most affected by spectral slope. Compared with the higher-scoring groups, they have the highest second peak higher response rates in the breathier-tenser condition, and the lowest response rates for tenser-breathier. The shift between the spectral slope conditions is smaller for the second and third quantiles, and the highest scoring groups four and five show the least amount of shift. It appears that group 5 has slightly more shift than group 4. This might be due to the fact that there are more subjects in group 5, and therefore there is a slightly larger individual variation in perceptual strategies than in group 4. A second observation from Fig. 6 is that the groups differ in the categoricity of their responses. Normally, categorical responses in a continuum result in a S-like curve, and this is indeed the case for groups with higher musicality scores. On the other hand, the response curves are more flat for groups with lower scores and are therefore less categorical. Overall, these results indicate that subjects with better musicality are less affected by spectral slope differences and more categorical in their f0 perception.

For each subject, a perceptual shift score and a categoricity score are calculated using methods outlined in Sec. II B. Both perceptual shift and categoricity scores show significant correlations with the subtest scores from the musicality experiment ($p < 0.05$). However, the correlations with the subtest scores are weaker than the correlations with the overall musicality scores. Therefore, the overall musicality scores are used in the analysis in this section.

The relationship between musicality and perceptual shift is illustrated in Fig. 7. There is a trend that subjects with higher musicality scores exhibit less shift as the result of spectral slope conditions. The correlation between musicality scores and perceptual shift is −0.3695 ($t = -4.2091$, $df = 112$, p-value < 0.001). A multiple regression is calculated to predict perceptual shift scores based on musicality scores, native language, and their interaction (shift∼musicality*language). The model explains a significant amount of the variance in perceptual shift [$F(3, 110) = 6.221$, $R^2 = 0.1451$, $R^2_{adjusted} = 0.1217$, $p = 0.0006$]. While musicality scores contribute significantly

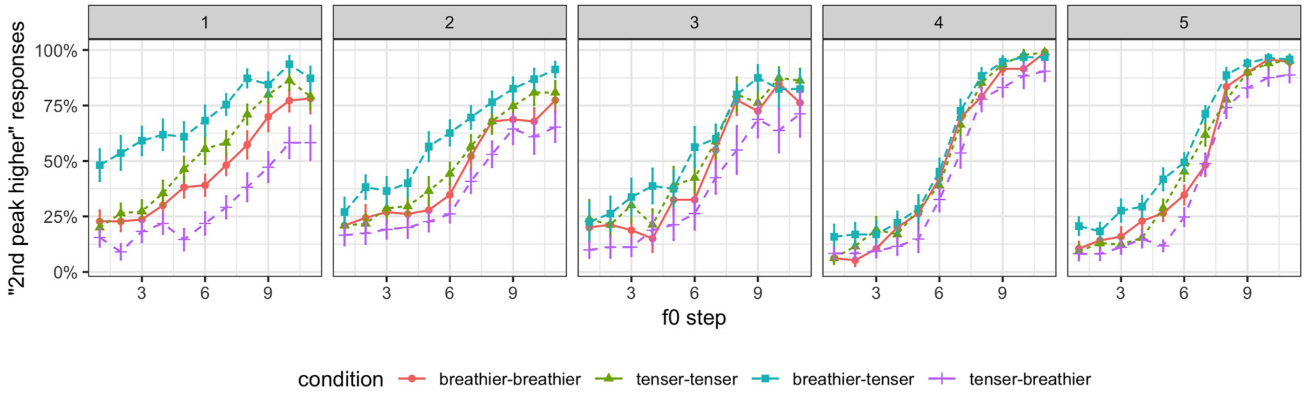

FIG. 6. (Color online) The subjects are divided into five quantiles by their total scores on the musicality test. The overall responses from the pitch perception experiment are plotted for each group, from lowest to highest.

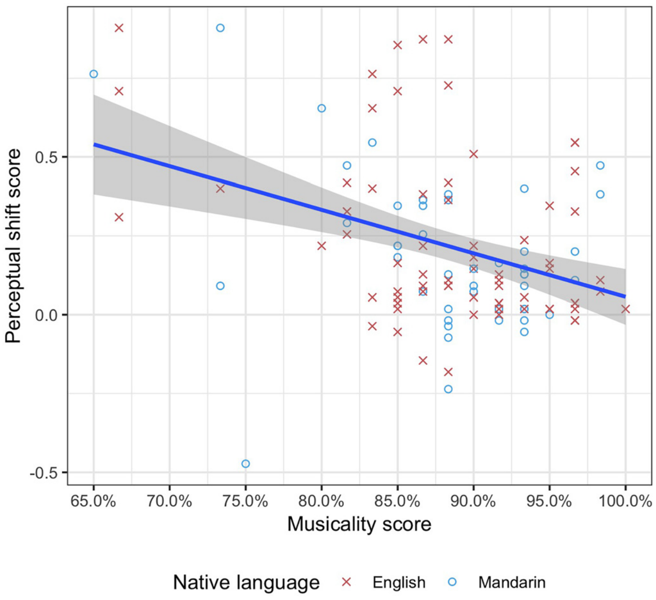

FIG. 7. (Color online) The relationship between musicality scores and perceptual shift scores for all the subjects. Shaded area = 95% confidence interval.

to the model ($\beta = -1.4814$, $p = 0.0006$), native language and its interaction with musicality scores do not.

Figure 8 plots the relationship between musicality scores and categoricity of the subject responses. Subjects with higher musicality scores show more categorical responses. The correlation between musicality scores and categoricity of perception is 0.4314 ($t = 5.0616$, $df = 112$, p-value < 0.001). A multiple regression is calculated to predict the categoricity scores based on musicality scores, native language, and their interaction (categoricity∼musicality*language). The model explains a significant amount of the variance in categoricity [$F(3, 110) = 8.716$, $R^2 = 0.1921$, $R^2_{adjusted} = 0.17$, $p < 0.0001$].

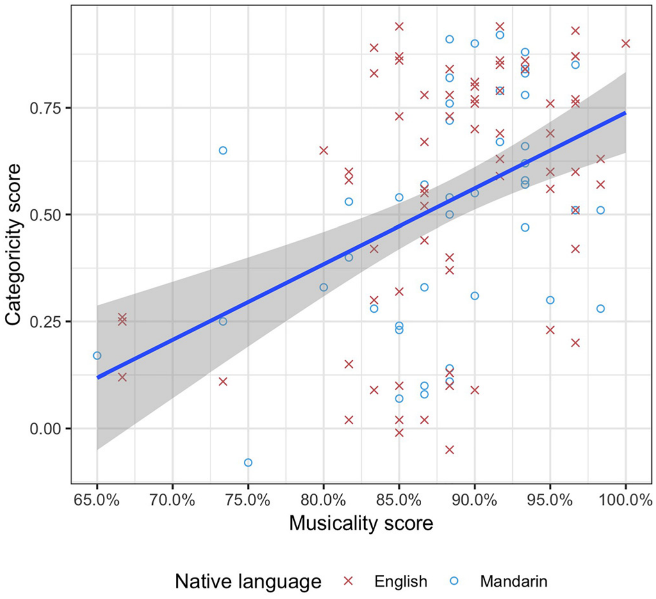

FIG. 8. (Color online) The relationship between musicality scores and categorical perception scores for all the subjects. Shaded area = 95% confidence interval.

Similar to the model for perceptual shift, while musicality scores contribute significantly to the model ($\beta = 1.771$, $p = 0.0001$), native language and its interaction with musicality scores do not.

Taken together, Figs. 7 and 8 demonstrate that subjects who have better musicality scores exhibit both smaller perceptual shift based on spectral conditions and greater categoricity along the f0 continuum. Therefore, the pattern observed in Fig. 6 is confirmed by these quantitative analyses.

To clearly address the methodological issue of whether musicianship or musicality scores would be a better predictor of pitch perception strategies, Fig. 9 is plotted to compare the effects of musicianship and musicality. In this plot, the perception by musicians and non-musicians (upper rows vs lower rows) are divided by the overall median (88.33%) musicality scores (left column vs right column). For both musicians and non-musicians, there is a clear difference between subjects who scored above the median and below the median. Regardless of musicianship, subjects with better musicality scores show smaller shift, whereas subjects with poorer musicality show greater shift. This plot demonstrates that musicality score is a better predictor for individual variation in pitch perception than musicianship.

IV. GENERAL DISCUSSION

This study investigates the relationship between musicality, tone language background, and pitch perception strategies by examining the correlations between musicality and pitch perception by both Mandarin and English speakers.

A. Musicality and pitch processing

The most important finding of this study is that musicality is a significant predictor of pitch processing strategies. Generally, better musicality indicates enhanced sensitivity to f0 and a preference for f0 cues in pitch perception. As shown in Figs. 7 and 8, even though all the subjects have relatively high scores on the musicality test, there is observable variation between subjects with lower scores and subjects with higher scores. Individual variation in pitch perception responses is quantified in two ways: the magnitude of shift between the spectral slope conditions and categoricity along the f0 continuum. Musicality scores correlate negatively with the magnitude of shift. This indicates that subjects with higher musicality scores are less affected by spectral slope differences and are more likely to judge relative pitch height based on f0 cues, while subjects with lower musicality scores are more likely to attend to spectral slope differences. These results are consistent with the previous study (Kuang and Liberman, 2015) and are also similar to Seither-Preisler *et al.* (2007), in which non-musicians primarily attend to spectral cues more and musicians primarily attend to f0 cues. A second result is that subjects with higher musicality scores have more categorical judgment along the f0 dimension, suggesting that they are more sensitive to f0 differences. This is also consistent with the previous findings that subjects with better musicality show greater categorical perception of f0 (Peng *et al.*, 2010).

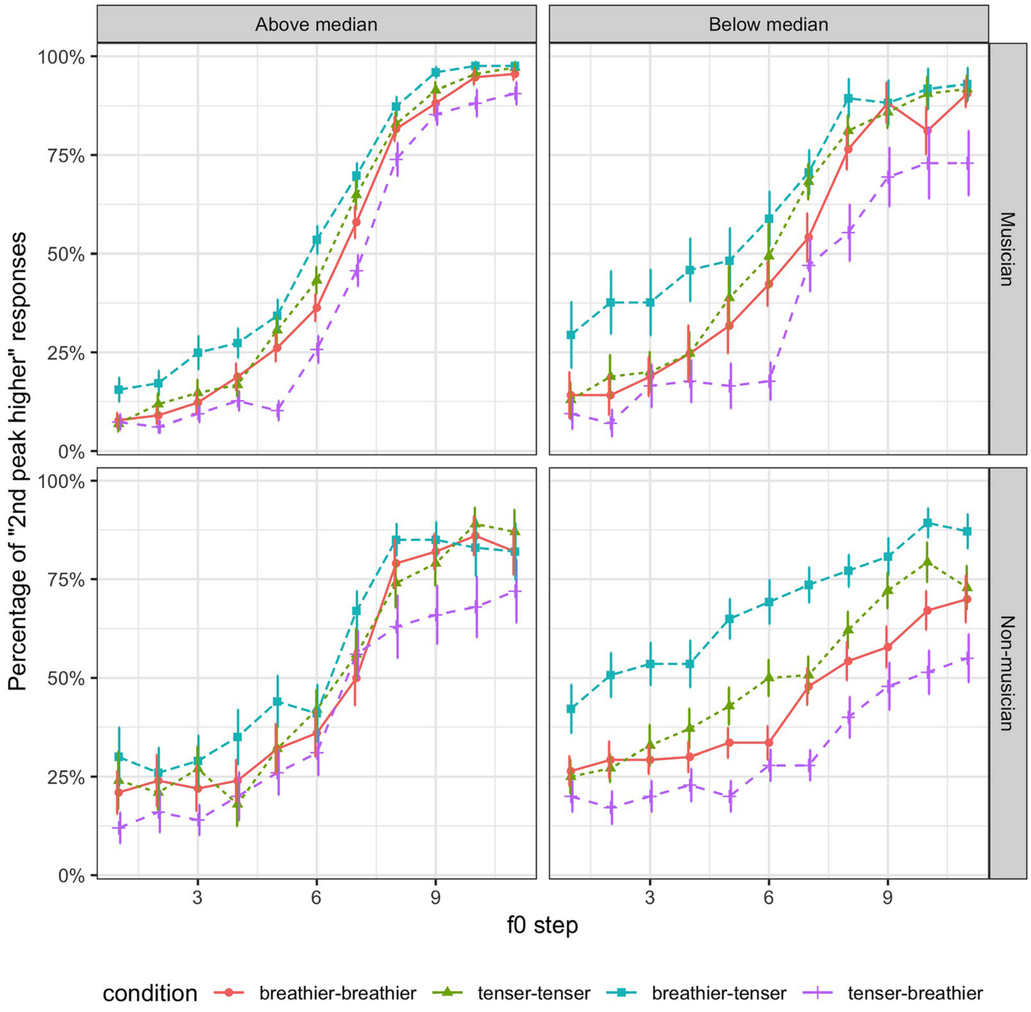

FIG. 9. (Color online) Results from the pitch perception experiment for musicians and non-musicians who scored above and below the overall group median on the musicality test.

One important methodological choice in this study is obtaining musicality scores from the administration of MBEMA (Peretz *et al.*, 2013) rather than only using self-identified musician labels. The overall mean musicality score 88.33% closely matches the mean (88.8%) found for adults in Peretz *et al.* (2013) for the abbreviated MBEMA. Even though subjects with music training generally have higher musicality scores, both musicians and non-musicians have scores that span the whole range of scores. Overall, these scores offer a continuous, more objective measure of musical aptitude than the binary distinction of musicianship. The musicality test makes it possible to identify non-musicians who are naturally more musically talented as well as musicians who still struggle with some aspects of music. Non-musicians with high musicality scores have similar patterns in the pitch perception experiment as high-scoring musicians, indicating that musicality is a better predictor of cue sensitivity than musicianship.

Moreover, these findings on musicality effects on pitch perception have important implications for language acquisition. The linguistic use of pitch involves multiple cues, and the pitch perception experiment shows that listeners vary in their use of f0 and spectral cues in relative pitch judgment. Therefore, speakers who experience difficulty in discriminating f0 can rely on additional spectral cues in acquiring linguistic distinctions. Although our subject pool mostly consists of people with good musical aptitude, there is a definite trend in cue integration in pitch judgment. Since less musical subjects tend to rely on spectral cues to make their judgment of pitch, it is possible that this trend will extend to amusic subjects as well. The results of this study suggest that speakers with amusia or tone agnosia may acquire normal tone production by relying on cues that co-vary with f0. Of course, this hypothesis needs to be further validated with data from children and the amusic population.

B. Language background and pitch processing

The second central research question of this study is whether language background has an effect on pitch processing. Since Mandarin speakers use pitch to indicate lexical tone contrast, it is possible that their pitch processing might differ from English speakers. Does speaking a tone language lead to better musicality? It appears not to be the case. Both the English and Mandarin groups have very similar musicality scores, indicating that experience with pitch processing in tonal contrasts does not grant speakers an advantage in music. More importantly, since there are multiple acoustic cues for pitch, does speaking a tone language have an effect on cue integration strategies in pitch perception? Again, language experience appears to have little effect. Both English and Mandarin speakers integrate f0 and spectral slope to perceive relative pitch height, and tenser voice sounds higher to both groups of speakers. Moreover, the magnitude of cue integration (as quantified in Figs. 7 and 8) is also similar in both languages.

Overall, these findings are consistent with the notion that language experience has little effect on general pitch perception (Ngo *et al.*, 2016; Zheng and Samuel, 2018). Although this study does not show any language effects, several notes of caution should be made. First, as discussed earlier in the paper, different studies have used different experimental paradigms for pitch perception. In particular, effects of language experience are more often reported when the tasks in question use speech stimuli rather than non-speech stimuli (Peng *et al.*, 2010; Xu *et al.*, 2006). However, it looks like the domain of the language effects is perhaps even more specific than speech vs non-speech. The stimuli in the current study are speech stimuli, but the pitch variation is more similar to intonation at the sentence level. It is possible that tone language experience might show effects specific to tasks related to lexical contrasts. In future studies, pitch tasks at different linguistic levels (e.g., tone vs intonation) should be tested. Second, the lack of language effect between Mandarin and English cannot be used to entirely rule out potential language effect from other tone languages. Mandarin tones mainly contrast in pitch direction (level, rising, falling, and dipping), but some other tone languages, such as Cantonese, have multiple contrastive level tones. It is possible that speakers of these tone languages might have enhanced general pitch perception. For example, although Bidelman *et al.* (2011) found that Mandarin speakers did not perform better in pitch discrimination than English speakers, Bidelman *et al.* (2013) showed enhancement for music perception in the Cantonese group. However, DiCanio (2012) found that Trique speakers (a four-level tone language) did not have such advantages. The effect of tone language background on general pitch perception therefore remains an open question for future studies.

V. CONCLUSION

This study investigates the effect of musicality and native language on pitch perception. Musicality is a significant predictor for cue integration strategies in pitch perception at the individual level. For relative pitch judgment, listeners with better musicality are more likely to attend to f0 cues, while listeners with poorer musicality are more likely to attend to spectral slope, the cue co-varying with f0. Additionally, at the group level, language experience has no effect on musicality or cue integration in pitch processing. These findings are generally consistent with the view that linguistic experience with pitch only results in domain-specific enhancement in pitch perception. Future work needs to extend this study to more linguistic tasks and other tone languages. Overall, this study sheds light on the complicated relationship between general pitch perception and linguistic pitch processing and has important implications for tone acquisition.

ACKNOWLEDGMENTS

We would like to thank Jacob LaViolette, Rhosean Asmah, and Leila Pearlman for their help with data collection. We would also like to thank Zhaoyan Zhang, Fan-Gang Zeng, and an anonymous reviewer for their insightful comments and suggestions on earlier versions of this paper.

Alexander, J. A., Wong, P. C., and Bradlow, A. R. (**2005**). "Lexical tone perception in musicians and non-musicians," in *Ninth European Conference on Speech Communication and Technology*, Lisbon, Portugal.

Allen, E. J., and Oxenham, A. J. (**2014**). "Symmetric interactions and interference between pitch and timbre," J. Acoust. Soc. Am. **135**(3), 1371–1379.

Baharloo, S., Service, S. K., Risch, N., Gitschier, J., and Freimer, N. B. (**2000**). "Familial aggregation of absolute pitch," Am. J. Hum. Genet. **67**(3), 755–758.

Bent, T., Bradlow, A. R., and Wright, B. A. (**2006**). "The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds," J. Exp. Psychol.: Hum. Percept. Perform. **32**(1), 97–103.

Best, C. T., Morrongiello, B., and Robson, R. (**1981**). "Perceptual equivalence of acoustic cues in speech and nonspeech perception," Percept. Psychophys. **29**(3), 191–211.

Bever, T. G. (**1975**). "Cerebral asymmetries in humans are due to the differentiation of two incompatible processes: Holistic and analytic," Ann. N. Y. Acad. Sci. **263**(1), 251–262.

Bidelman, G. M., Gandour, J. T., and Krishnan, A. (**2011**). "Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch," Brain Cognit. **77**(1), 1–10.

Bidelman, G. M., 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**(4), e60676.

Chen, A., Liu, L., and Kager, R. (**2016**). "Cross-domain correlation in pitch perception, the influence of native language," Lang., Cognit. Neurosci. **31**(6), 751–760.

Chen, A., Peter, V., Wijnen, F., Schnack, H., and Burnham, D. (**2018**). "Are lexical tones musical? native language's influence on neural response to pitch in different domains," Brain Lang. **180**, 31–41.

Chin, C. S. (**2003**). "The development of absolute pitch: A theory concerning the roles of music training at an early developmental age and individual cognitive style," Psychol. Music **31**(2), 155–171.

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**(4), 2398–2403.

Deutsch, D., Henthorn, T., Marvin, E., and Xu, H. (**2006**). "Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period," J. Acoust. Soc. Am. **119**(2), 719–722.

DiCanio, C. T. (**2012**). "Cross-linguistic perception of itunyoso trique tone," J. Phonet. **40**(5), 672–688.

Escudero, P., Benders, T., and Lipski, S. C. (**2009**). "Native, non-native and l2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners," J. Phonet. **37**(4), 452–465.

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. Phonet. **36**(2), 268–294.

François, C., Chobert, J., Besson, M., and Schön, D. (**2012**). "Music training for the development of speech segmentation," Cereb. Cortex **23**(9), 2038–2043.

Fujioka, T., Ross, B., Kakigi, R., Pantev, C., and Trainor, L. J. (**2006**). "One year of musical training affects development of auditory cortical-evoked fields in young children," Brain **129**(10), 2593–2608.

Gandour, J. (**1983**). "Tone perception in far eastern languages," J. Phonet. **11**, 149–178.

Gandour, J. T., and Harshman, R. A. (**1978**). "Crosslanguage differences in tone perception: A multidimensional scaling investigation," Lang. Speech **21**(1), 1–33.

Gobl, C., and Ní Chasaide, A. (**2010**). "Voice source variation and its communicative functions," in *The Handbook of Phonetic Sciences*, 2nd ed., edited by W. J. Hardcastle, J. Laver, and F. E. Gibbon (Blackwell, Oxford), pp. 378–423.

Gregersen, P. K., Kowalsky, E., Kohn, N., and Marvin, E. W. (**2001**). "Early childhood music education and predisposition to absolute pitch: Teasing apart genes and environment," Am. J. Med. Genet. **98**(3), 280–282.

Hadfield, J. D. (**2010**). "MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package," J. Stat. Software **33**(2), 1–22, http://www.jstatsoft.org/v33/i02/.

Holt, L. L., and Lotto, A. J. (**2006**). "Cue weighting in auditory categorization: Implications for first and second language acquisition," J. Acoust. Soc. Am. **119**(5), 3059–3071.

Hutchins, S., and Moreno, S. (**2013**). "The linked dual representation model of vocal perception and production," Front. Psychol. **4**, 825.

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.

Kalmus, H., and Fry, D. (**1980**). "On tune deafness (dysmelodia): Frequency, development, genetics and musical background," Ann. Hum. Genet. **43**(4), 369–382.

Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., and Banno, H. (**2008**). "Tandem-straight: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, f0, and aperiodicity estimation," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, New York, pp. 3933–3936.

Koelsch, S., Gunter, T. C., Cramon, D. Y. v., Zysset, S., Lohmann, G., and Friederici, A. D. (**2002**). "Bach speaks: A cortical 'language-network' serves the processing of music," Neuroimage **17**(2), 956–966.

Kong, Y.-Y., and Zeng, F.-G. (**2006**). "Temporal and spectral cues in mandarin tone recognition," J. Acoust. Soc. Am. **120**(5), 2830–2840.

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

Krumhansl, C. L., and Iverson, P. (**1992**). "Perceptual interactions between musical pitch and timbre," J. Exp. Psychol.: Hum. Percept. Perform. **18**(3), 739–751.

Kuang, J. (**2017**). "Covariation between voice quality and pitch: Revisiting the case of mandarin creaky voice," J. Acoust. Soc. Am. **142**(3), 1693–1706.

Kuang, J., and Liberman, M. (**2015**). "The effect of spectral slope on pitch perception," in *Sixteenth Annual Conference of the International Speech Communication Association*.

Kuang, J., and Liberman, M. (**2016**). "The effect of vocal fry on pitch perception," in *Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5260–5264.

Kuang, J., and Liberman, M. (**2018**). "Integrating voice quality cues in the pitch perception of speech and non-speech utterances," Front. Psychol. **9**, 2147.

Kuhl, P. K., and Miller, J. D. (**1978**). "Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli," J. Acoust. Soc. Am. **63**(3), 905–917.

Ladd, D. R., Turnbull, R., Browne, C., Caldwell-Harris, C., Ganushchak, L., Swoboda, K., Woodfield, V., and Dediu, D. (**2013**). "Patterns of individual differences in the perception of missing-fundamental tones," J. Exp. Psychol.: Hum. Percept. Perform. **39**(5), 1386–1397.

Law, L. N., and Zentner, M. (**2012**). "Assessing musical abilities objectively: Construction and validation of the profile of music perception skills," PLoS One **7**(12), e52508.

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

Lipski, S. C., Escudero, P., and Benders, T. (**2012**). "Language experience modulates weighting of acoustic cues for vowel perception: An event-related potential study," Psychophysiology **49**(5), 638–650.

Liu, F., Chan, A. H., Ciocca, V., Roquet, C., Peretz, I., and Wong, P. C. (**2016**). "Pitch perception and production in congenital amusia: Evidence from Cantonese speakers," J. Acoust. Soc. Am. **140**(1), 563–575.

Magne, C., Schön, D., and Besson, M. (**2006**). "Musician children detect pitch violations in both music and language better than nonmusician children: Behavioral and electrophysiological approaches," J. Cognit. Neurosci. **18**(2), 199–211.

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. Cognit. Neurosci. **23**(10), 2701–2715.

Marozeau, J., and de Cheveigné, A. (**2007**). "The effect of fundamental frequency on the brightness dimension of timbre," J. Acoust. Soc. Am. **121**(1), 383–387.

Melara, R. D., and Marks, L. E. (**1990**). "Interaction among auditory dimensions: Timbre, pitch, and loudness," Percept. Psychophys. **48**(2), 169–178.

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., and Besson, M. (**2009**). "Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity," Cereb. Cortex **19**(3), 712–723.

Nan, Y., Sun, Y., and Peretz, I. (**2010**). "Congenital amusia in speakers of a tone language: Association with lexical tone agnosia," Brain **133**(9), 2635–2642.

Neuhoff, J. G., and McBeath, M. K. (**1996**). "The Doppler illusion: The influence of dynamic intensity change on perceived pitch," J. Exp. Psychol.: Human Percept. Perform. **22**(4), 970–985.

Ngo, M. K., Vu, K.-P. L., and Strybel, T. Z. (**2016**). "Effects of music and tonal language experience on relative pitch performance," Am. J. Psychol. **129**(2), 125–134.

Patel, A. D., Gibson, E., Ratner, J., Besson, M., and Holcomb, P. J. (**1998**). "Processing syntactic relations in language and music: An event-related potential study," J. Cognit. Neurosci. **10**(6), 717–733.

Patel, A. D., Wong, M., Foxton, J., Lochy, A., and Peretz, I. (**2008**). "Speech intonation perception deficits in musical tone deafness (congenital amusia)," Music Percept.: Interdisciplinary J. **25**(4), 357–368.

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. Phonet. **38**(4), 616–624.

Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B., and Jutras, B. (**2002**). "Congenital amusia: A disorder of fine-grained pitch discrimination," Neuron **33**(2), 185–191.

Peretz, I., Cummings, S., and Dubé, M.-P. (**2007**). "The genetics of congenital amusia (tone deafness): A family-aggregation study," Am. J. Hum. Genet. **81**(3), 582–588.

Peretz, I., Gosselin, N., Nan, Y., Caron-Caplette, E., Trehub, S. E., and Béland, R. (**2013**). "A novel tool for evaluating children's musical abilities across age and culture," Front. Syst. Neurosci. **7**, 30.

Peretz, I., Vuvan, D., Lagrois, M.-É., and Armony, J. L. (**2015**). "Neural overlap in processing music and speech," Philos. Trans. R. Soc., London B **370**(1664), 20140090.

Pfordresher, P. Q., and Brown, S. (**2009**). "Enhanced production and perception of musical pitch in tone language speakers," Atten., Percept., Psychophys. **71**(6), 1385–1398.

Posedel, J., Emery, L., Souza, B., and Fountain, C. (**2012**). "Pitch perception, working memory, and second-language phonological production," Psychol. Music **40**(4), 508–517.

Schön, 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.

Seither-Preisler, A., Johnson, L., Krumbholz, K., Nobbe, A., Patterson, R., Seither, S., and Lütkenhöner, B. (**2007**). "Tone sequences with conflicting fundamental pitch and timbre changes are heard differently by musicians and nonmusicians," J. Exp. Psychol. **33**(3), 743.

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**(2), 273–293.

Stevens, C. J., Keller, P. E., and Tyler, M. D. (**2013**). "Tonal language background and detecting pitch contour in spoken and musical items," Psychol. Music **41**(1), 59–74.

Titze, I. R. (**1988**). "A framework for the study of vocal registers," J. Voice **2**(3), 183–194.

Wayland, R. P., and Guion, S. G. (**2004**). "Training English and Chinese listeners to perceive Thai tones: A preliminary report," Lang. Learn. **54**(4), 681–712.

Wong, P. C., and Perrachione, T. K. (**2007**). "Learning pitch patterns in lexical identification by native English-speaking adults," Appl. Psycholinguist. **28**(4), 565–585.

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**(2), 1063–1074.

Zatorre, R. J., Belin, P., and Penhune, V. B. (**2002**). "Structure and function of auditory cortex: Music and speech," Trends Cognit. Sci. **6**(1), 37–46.

Zhang, Z. (**2016**). "Mechanics of human voice production and control," J. Acoust. Soc. Am. **140**(4), 2614–2635.

Zheng, Y., and Samuel, A. G. (**2018**). "The effects of ethnicity, musicianship, and tone language experience on pitch perception," Q. J. Exp. Psychol. **71**(12), 2627–2642.