Perception of pitch height in lexical and musical tones by English-speaking musicians and nonmusicians

Chao-Yang Lee, Allison Lekich, and Yu Zhang
Division of Communication Sciences and Disorders, Ohio University, Athens, Ohio 45701

(Received 31 May 2013; revised 7 January 2014; accepted 24 January 2014)

The purpose of this study was to explore the music-speech relationship by examining pitch height perception in lexical and musical tones. English-speaking musicians and nonmusicians identified multispeaker Taiwanese level tones without typical cues for speaker normalization. The musicians also identified note names of piano, viola, and pure tones without a reference pitch. In the Taiwanese task, both the musicians and nonmusicians were able to identify tone height above chance, but only for tones at the extremes of the speakers’ overall vocal range. The musicians only had a slight advantage over the nonmusicians. In the music task, none of the musicians met the criterion for absolute pitch. Timbre did not affect how accurately the musical tones were identified. No correlations were found between performance in the Taiwanese task and that in the music task. It was concluded that musicians’ advantage in lexical tone perception arose from the ability to track F0 contours. The ability to identify pitch height in lexical tones appears to involve calibrating acoustic input according to gender-specific, internally stored pitch templates.

© 2014 Acoustical Society of America. [http://dx.doi.org/10.1121/1.4864473]

PACS number(s): 43.75.Cd, 43.71.Hw, 43.71.Bp [DD] Pages: 1607–1615

I. INTRODUCTION

Pitch is used in many tone languages to distinguish between segmentally identical words. Pitch is also a major attribute of music. Since pitch is an integral part of lexical and musical tones, an intriguing question is whether similar cognitive mechanisms underlie lexical and musical pitch perception. The relationship between music and speech/language has attracted interest from scholars from diverse disciplines (Patel, 2008). The nature of the relationship, however, remains widely debated. The goal of this study was to explore the relationship by examining pitch height perception in lexical and musical tone identification. In particular, we examined the identification of lexical tones by English-speaking musicians and nonmusicians. We also investigated the musicians’ ability to identify note names of musical tones without a reference pitch (Deutsch et al., 2006). Results from these two experiments would be synthesized with findings from our previous studies (Lee and Hung, 2008; Lee and Lee, 2010; Lee et al., 2011) to explicate the roles of tone language experience and musical training in lexical and musical pitch perception.

The link between music and speech/language processing is supported by studies showing that musical training facilitates nontone language speakers’ ability to produce and perceive lexical tones (Alexander et al., 2005; Gottfried, 2007; Gottfried and Riester, 2000; Gottfried et al., 2001). However, why musical training facilitates lexical pitch processing is not fully understood. Deutsch and colleagues (Deutsch, 2002, 2006, 2013; Deutsch et al., 1999, 2004; Deutsch et al., 2006; Deutsch et al., 2009) proposed that lexical tone is associated with absolute pitch, i.e., the ability to produce or name a musical tone without a reference pitch. Since absolute pitch and lexical tones both involve associating pitch with a verbal label, the idea is that absolute pitch and lexical tones share a common evolutionary origin. Based on this proposal, one would expect absolute pitch to be more prevalent in tone language speakers. One would also expect tone language speakers to possess more stable and precise pitch templates. Finally, one would expect the acquisition of absolute pitch to resemble acquisition of speech in that prevalence of absolute pitch would be negatively correlated with age of onset of musical training. These predictions are supported by studies on speaking F0 consistency (Deutsch et al., 1999, 2004) and relative prevalence of absolute pitch (Deutsch et al., 2006; Deutsch et al., 2009) in tone and nontone language speakers.

In addition to the research on speaking F0 and prevalence of absolute pitch, a series of studies by Lee and colleagues further explored the perception of lexical and musical pitch as a function of tone language experience and musical training. Lee and Hung (2008) examined Mandarin tone identification by 36 English-speaking musicians and 36 nonmusicians. Mandarin /sa/ syllables, produced with the four lexical tones (high-level, mid-rising, mid-dipping, and high-falling) by 16 female and 16 male speakers, were digitally edited to generate intact syllables, silent-center syllables (in which 70% of the syllable center was removed; only the onset consonant, beginning 15%, and ending 15% of the voiced portion of the syllable remained in the stimuli) and onset-only syllables (in which only the onset consonant and beginning 15% of the voiced portion of the syllable remained in the stimuli). Truncated syllables have been widely used in speech perception studies to assess how well listeners could use limited acoustic information to retrieve
vowel or tone identity (e.g., Strange, 1989; Gottfried and Suiter, 1997). The results showed that the musicians outperformed the nonmusicians, although the advantage diminished as the amount of F0 information was reduced. In particular, the musicians’ advantage was 24% for intact syllables, 18% for silent-center syllables, and 3% for onset-only syllables. Importantly, some of the onset-only tones could be identified even though they were quite brief, presented in isolation, and without repetition, suggesting that the listeners were able to use syllable–internal pitch height to identify the tones despite lacking typical cues for speaker normalization, e.g., F0 contour, external context, and familiarity with speakers (Leather, 1983; Moore and Jongman, 1997; Johnson, 2005).

In addition to the Mandarin task, the musicians were also given a music task to evaluate their ability to identify musical tones without a reference pitch. Three sets of synthesized piano, viola, and pure tones were presented to the listeners. These 500 ms long tones ranged from C3 to B4 in pitch. Consecutive tones were always separated by more than one octave to prevent participants from using relative pitch for the task. The participants responded by notating on staff paper the tones that they had heard. The results showed that none of the musicians met the criterion for absolute pitch, defined as 85% or higher accuracy in identifying the piano tones (Deutsch et al., 2006). Correlations were calculated between performance in Mandarin tone identification and musical note identification to evaluate the association between lexical and musical pitch perception. The results showed that none of the correlations were statistically significant. However, since none of the musician participants in the study actually possessed absolute pitch, it was not clear whether a significant correlation would emerge for musicians with absolute pitch.

Higher occurrence of absolute pitch has been reported in musicians speaking a tone language (Deutsch et al., 2004; Deutsch et al., 2006). Therefore, Lee and Lee (2010) examined Mandarin tone and musical tone identification by 72 Mandarin-speaking musicians. In the Mandarin task, the onset-only /sa/ syllables used in Lee and Hung (2008) were presented to the listeners. Intact and silent-center syllables were not included because previous studies had shown that native Mandarin listeners could identify tones from those syllables very well (Gottfried and Suiter, 1997; Lee et al., 2008). More importantly, the lack of distinct F0 contours in the onset-only tones would be particularly useful in probing the use of pitch height in lexical tone identification. The results showed that the Mandarin-speaking musicians could identify the majority of the onset-only tones above chance. A similar result was found in an earlier study with Mandarin-speaking nonmusicians (Lee, 2009), suggesting native listeners could identify pitch height without typical cues of speaker normalization, irrespective of their musical training. In the music task, 72% of the Mandarin-speaking musicians met the criterion for absolute pitch, which contrasted sharply with the absence of absolute pitch possessors in the English speakers in Lee and Hung (2008). However, there were still no significant correlations between Mandarin tone and musical tone identification, whether all musicians or just those with absolute pitch were included in the analysis. In other words, even with a substantial sample of absolute pitch possessors, there was still no evidence for the association between lexical and musical tone perception.

Although onset-only Mandarin tones were analogous to the stimuli used in the musical tone task in that they included only a single, flat F0, the mental representation of Mandarin tones is still contour in nature. That means the contrast among the four Mandarin tones could be processed based on F0 contour, without reference to pitch height. Our next goal, therefore, was to identify a language with canonical level tones in which pitch height has to be used in order to identify the tones. To that end, Lee et al. (2011) identified two pairs of Taiwanese words contrasting only in pitch height. The syllables /hue/ (“flower” with a high tone, and “meeting” with a mid tone) and /di/ (“pig” with a high tone, and “chopsticks” with a mid tone) were recorded by 15 female and 15 male native speakers. The stimuli were presented in isolation to 43 Taiwanese-speaking musicians. The results showed that the participants could identify tone height above chance without typical cues for speaker normalization. In the music task, 65% of the participants met the criterion for absolute pitch. However, there were still no significant correlations between performance in the lexical tone task and that in the musical tone task.

This series of studies revealed several patterns regarding pitch perception in tone language and music. First, the percentage of absolute pitch possessors among the musicians was substantially higher in speakers of a tone language. There was also a negative correlation between performance in the music task and age of onset of musical training. These two findings are consistent with Deutsch’s proposal that absolute pitch is associated with lexical tones. Second, among the nontone language (English-speaking) participants, musicians had an advantage over nonmusicians in lexical tone identification. However, the advantage diminished as the amount of acoustic information was reduced. The musicians’ advantage was also smaller for level tones compared to contour tones. These two results, along with the finding that none of the English-speaking musicians possessed absolute pitch, suggest that the ability to track pitch contours, rather than pitch height, is likely the source of the musicians’ advantage.

Although this series of studies allowed us to systematically evaluate the roles of tone language experience and musical training in lexical and musical pitch perception, how well nontone language speakers could identify canonically level tones in language (e.g., Taiwanese level tones) remains unknown. This information is critical in evaluating the role of pitch height perception in lexical tone identification. To supply the missing piece, two experiments were conducted in this study to examine Taiwanese level tone identification and musical tone identification by English-speaking musicians and nonmusicians. In the Taiwanese task, both musicians and nonmusicians were asked to identify Taiwanese level tones, which involved judging vocal pitch height. In the music task, the musicians were asked to identify musical note names without a reference pitch, which involved judging musical pitch height. Our goal was to compare the results
to previous studies using similar tasks (Lee and Hung, 2008; Lee and Lee, 2010; Lee et al., 2011) to further elucidate the relationship between lexical and musical pitch perception.

II. EXPERIMENT 1: TAIWANESE TONE IDENTIFICATION

In this experiment, English-speaking musicians and nonmusicians were asked to identify the pitch height of isolated Taiwanese level tones produced by multiple speakers. Tonal minimal pairs that contrast only in F0 height (high vs mid) were chosen as stimuli. Because F0 range varies across speakers, F0 height judgment had to be made with reference to a speaker’s F0 range. In this experiment, acoustic cues that are typically present for speaker normalization were removed or minimized by presenting level-tone stimuli (no F0 contour) produced in isolation (no context) by speakers whose voices were never heard by listeners (no familiarity). This allowed for an evaluation of contribution of syllable-internal F0 information to pitch height perception. If musical training facilitates lexical pitch perception, one would expect musicians to identify the level tones more accurately than nonmusicians. Furthermore, if participants are able to use pitch height to identify the level tones, one would expect identification accuracy to exceed chance.

A. Method

1. Materials

The materials used in this experiment were identical to the Taiwanese-tone stimuli used in Lee et al. (2011). They include two pairs of Taiwanese words, /hue/ and /hue/ with a high tone, “flower”) and /hue/ and /hue/ with a mid tone, “meeting”) as well as /di/ (“pig”) and /di/ (“chopsticks”). As the transcriptions indicate, members within a pair contrast minimally in tone, and the tonal contrast involves pitch height only. These 4 words were recorded by 30 adult speakers of Taiwanese (15 females and 15 males). Details of the recording procedure and acoustic analysis are described in Lee et al. (2011). Figure 1, adapted from Lee et al. (2011), shows the F0 contours of the stimuli. The average F0s for the stimuli are shown in Table I. The acoustic data show that female speakers have overall higher F0s than male speakers. In general, the female and male speakers have distinct F0 ranges, indicated by the fact that mid tones produced by the female speakers are generally higher than high tones produced by the male speakers. This observation suggests that pitch height judgment could be facilitated by successful identification of speaker gender (Lee et al., 2010). By contrast, within a gender group there is substantial overlap between the high and mid tones, suggesting that tone identification could still be challenging even if speaker gender is known. Finally, the difference between the high and mid tones is greater for the female (40 Hz) than male (27 Hz) speakers, suggesting that tones produced by female speakers may be easier to identify. In terms of semitones, the high-mid tone difference is comparable between the female (3.06) and male (3.18) speakers.

2. Participants

Seventy-two students at Ohio University participated in the experiment. They were all native speakers of American English without knowledge of or substantial exposure to tone languages. All participants were screened for normal hearing, defined as pure-tone, air-conducted thresholds of ≤20 dB hearing level at octave frequencies from 1000 to
3. Procedure

The procedure was identical to the Taiwanese tone identification experiment reported in Lee et al. (2011). Since the goal was to evaluate how listeners identify tone height from syllable-internal F0 information alone, efforts were made to remove or minimize syllable-extrinsic acoustic cues that might be used for speaker normalization. To that end, the 120 stimuli were assigned to 4 blocks of 30 stimuli. Two blocks included the pair of /di/ words and two blocks included the pair of /hue/ words. The 30 stimuli in a given block were spoken by all 30 speakers such that no individual speaker was heard more than once. The number of male and female speakers in a block was balanced (15 female stimuli and 15 male stimuli). The number of high and mid tones in a block was also balanced (15 high-tone stimuli and 15 mid-tone stimuli).

Participants were tested individually in a sound-treated booth. Prior to the experiment, the experimenter explained the lexically contrastive function of tones in Taiwanese. The participants were told that they would be listening to the 4 Taiwanese words produced by 30 speakers that contrast in pitch height. Their task was to identify the height of the tones by pressing computer keys labeled with “HIGH” and “LOW.” Eight practice trials, recorded by a female speaker and a male speaker, were given to familiarize the participants with the experimental procedure and response format. The two speakers who recorded the practice trials were not used in the actual experiment.

The stimuli were presented with the AVRunner program of the Brown Lab Interactive Speech System (BLISS; Mertus, 2000). The participants listened to the stimuli through a pair of headphones (KOSS SB/45, Milwaukee, WI) connected to a Windows laptop computer (Toshiba Satellite L755-S5277, Irvine, CA). Each participant was assigned a uniquely randomized presentation order, such that no two participants received the same order of presentation. The order of the blocks was also randomized for each participant. The participants were told that they had 5 s to respond to each word. They were instructed to respond as quickly as possible because the responses would be timed.

### TABLE I. Average F0s (in Hz) of the stimuli used in the Taiwanese tone identification experiment.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Tone</th>
<th>Syllable</th>
<th>F0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>High</td>
<td>/di/</td>
<td>247 (23)</td>
</tr>
<tr>
<td></td>
<td>Mid</td>
<td>/di/</td>
<td>207 (19)</td>
</tr>
<tr>
<td>Male</td>
<td>High</td>
<td>/di/</td>
<td>164 (31)</td>
</tr>
<tr>
<td></td>
<td>Mid</td>
<td>/di/</td>
<td>135 (28)</td>
</tr>
</tbody>
</table>

4000 Hz. The 36 musicians included 22 females and 14 males. They were music majors with the following concentrations: cello (2 individuals), clarinet (2), double bass (1), euphonium (2), flute (2), horn (1), oboe (1), piano (6), saxophone (4), trombone (1), trumpet (6), tuba (2), viola (1), violin (3), and voice (2). The age of the musicians ranged from 18 to 25 years (M = 20.5, standard deviation (SD) = 2.12). The age at which they started musical training ranged from 4 to 14 years (M = 8.89, SD = 2.39). The 36 nonmusicians included 28 females and 8 males. The age of the nonmusicians ranged from 18 to 25 years (M = 20.67, SD = 1.2). None of the nonmusician participants reported any formal music training or substantial music learning experience.

### B. Results and discussion

Identification responses were automatically acquired by BLISS. Accuracy (i.e., percentage of correct responses) and reaction time are shown in Table II. Overall, accuracy appears to be slightly higher for the musicians. For both musicians and nonmusicians, responses to high-tone stimuli produced by the female speakers and low-tone stimuli produced by the male speakers appear to be more accurate and faster. That is, stimuli in the extremes of the speakers’ vocal F0 range seem easier to identify.

Before parametric tests were conducted to evaluate these observations, Kolmogorov–Smirnov tests and Bartlett’s tests were conducted to test the assumptions of normality and homogeneity of variance. A set of 16 Kolmogorov–Smirnov tests (2 speaker genders × 2 tone heights × 2 participant groups × 2 dependent variables) verified that distributions of both accuracy and reaction time data met the assumption of normal distribution. However, Bartlett’s tests showed that variances across the groups were not equal. Therefore, the accuracy data were arcsine-transformed and the reaction time data were log-transformed before parametric tests were conducted.

An analysis of variance (ANOVA) was conducted on arcsine-transformed accuracy with speaker gender (female and male) and tone height (high and mid) as within-subject factors, musical training (musicians and nonmusicians) as a between-subject factor, and participants as a random factor. The ANOVA showed a significant main effect of musical training, F(1,70) = 5.96, p = 0.017, ηp² = 0.078, a main effect of speaker gender, F(1,70) = 33.46, p < 0.001, ηp² = 0.323, and a significant speaker gender × tone height interaction, F(1,70) = 256.13, p < 0.001, ηp² = 0.785. The effect of musical training confirmed the observation that musicians identified tone height with higher accuracy than nonmusicians. The effect of speaker gender supported the prediction that responses to female stimuli would be more accurate.

---

TABLE II. Mean accuracy (in % with SD) and reaction time (in ms with SD) of Taiwanese tone identification. Also shown are results from t-tests evaluating whether the accuracy was above chance (50%).

<table>
<thead>
<tr>
<th>Listener</th>
<th>Speaker</th>
<th>Tone</th>
<th>Accuracy (pr)</th>
<th>r²</th>
<th>Reaction time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Musicians</td>
<td>Female</td>
<td>High</td>
<td>85 (11)</td>
<td>14.43</td>
<td>&lt;0.001 0.856</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mid</td>
<td>49 (18)</td>
<td>-0.38</td>
<td>0.704 0.004</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>High</td>
<td>44 (12)</td>
<td>-2.75</td>
<td>0.009 0.177</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mid</td>
<td>81 (7)</td>
<td>21.49</td>
<td>&lt;0.001 0.930</td>
</tr>
<tr>
<td>Nonmusicians</td>
<td>Female</td>
<td>High</td>
<td>83 (8)</td>
<td>18.68</td>
<td>&lt;0.001 0.909</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mid</td>
<td>46 (15)</td>
<td>-1.56</td>
<td>0.127 0.065</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>High</td>
<td>46 (12)</td>
<td>-2.94</td>
<td>0.049 0.106</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mid</td>
<td>76 (8)</td>
<td>16.92</td>
<td>&lt;0.001 0.891</td>
</tr>
</tbody>
</table>
The speaker gender × tone height interaction also confirmed the observation that success of tone height identification depends on speaker gender. That is, identification was more accurate when the stimuli were at the extremes of the speakers’ overall vocal range. For stimuli produced by the female speakers, the high tones were identified more accurately than the mid tones. For stimuli produced by the male speakers, the mid tones were identified more accurately than the high tones. These results were further supported by positive correlations between F0 and accuracy for the female stimuli (musicians: \( r = 0.56, p < 0.001 \); nonmusicians: \( r = 0.59, p < 0.001 \)) and negative correlations for the male stimuli, although the negative correlations were not statistically significant (musicians: \( r = -0.19, p = 0.15 \); nonmusicians: \( r = -0.11, p = 0.4 \)). Overall, the data from the English-speaking participants contrast with those from the native Taiwanese listeners (Lee et al., 2011) who showed highest accuracy for high-tone stimuli produced by the female speakers (76%) and comparable accuracy among the other three types of stimuli (58%, 61%, and 63% for low-female, high-male, and low-male stimuli, respectively). This contrast suggests that the ability to identify tone height is shaped by experience with a tone language.

To evaluate whether the accuracy of identification was above chance (50%), eight t-tests (2 speaker genders × 2 tone heights × 2 participant groups) were conducted on arcsine-transformed accuracy. The results, also shown in Table II, indicated that both musicians and nonmusicians were able to identify tone height above chance only for high-tone stimuli produced by the female speakers and low-tone stimuli produced by the male speakers. That is, only stimuli at the extremes of the speakers’ overall vocal range were identified above chance. By contrast, the listeners were not able to identify reliably the height of the tones produced in the middle of the speakers’ overall vocal range. Intuitively, this finding was to be expected because the stimuli in the middle of the range involve substantial overlap as shown in Fig. 1. However, this finding contrasts with performance of the native Taiwanese listeners (Lee et al., 2011), who were able to identify all groups of stimuli above chance. This contrast suggests that native Taiwanese listeners may be able to extract additional information from the acoustic signal or use alternative strategies to identify tone height.

Reaction time was measured from stimulus offset to avoid the potential confound of variable stimulus durations. Only correct responses were included in the reaction time analysis. An ANOVA was conducted on log-transformed reaction time with speaker gender (female and male) and tone height (high and mid) as within-subject factors, musical training (musicians and nonmusicians) as a between-subject factor, and participants as a random factor. The ANOVA showed a significant main effect of tone height, \( F(1,70) = 26.89, p < 0.001 \), \( \eta^2_p = 0.278 \), and a significant speaker gender × tone height interaction, \( F(1,70) = 121.08, p < 0.001 \), \( \eta^2_p = 0.634 \). Overall, the effect of tone height indicated that overall reaction time was longer for the mid tone than for the high tone. However, the higher-order speaker gender × tone height interaction further revealed that listeners took less time to respond to tones that are at the extremes of the speakers’ vocal F0 range. This result is consistent with the accuracy analysis, suggesting that tones produced at the extremes of the speakers’ overall vocal range were easier to identify.

In summary, results from this experiment showed that the musicians identified tone height better than the nonmusicians, although the advantage was minimal. For both groups of participants, responses to tone stimuli at the extremes of the speakers’ overall vocal range were more accurate and faster. In addition, identification accuracy exceeded chance only for tone stimuli at the extremes of the speakers’ overall range. This pattern contrasts with the native Taiwanese listeners’ performance reported in Lee et al. (2011).

III. EXPERIMENT 2: MUSICAL TONE IDENTIFICATION

The purpose of this experiment was to evaluate the musician participants’ ability to identify note names of musical tones without a reference pitch. Synthesized musical tones of three timbres were presented to the musicians whose task was to identify note names of the musical tones. Since no reference pitch was given, the listeners had to rely on their internal pitch scale to identify the tones. This task has been used in previous studies to examine the presence of absolute pitch (Deutsch et al., 2006; Lee and Hung, 2008; Deutsch et al., 2009; Lee and Lee, 2010; Lee et al., 2011).

A. Method

1. Materials

The materials used in this experiment were identical to those used in Lee and Hung (2008), Lee and Lee (2010), and Lee et al. (2011). Thirty-six notes that spanned a 3-octave range from C3 (131 Hz) to B5 (988 Hz) were synthesized with 3 timbres (piano, pure tone, and viola) for a total of 108 notes. The pure tones were synthesized with MATLAB (The MathWorks, Natick, MA). The piano and viola stimuli were synthesized with a Kurzweil K2000 synthesizer (North Andover, MA) tuned to the standard A4 at 440 Hz. The duration for all tones was 500 ms.

For each timbre, the 36 tones were ordered such that any 2 consecutive tones differed by more than an octave. The purpose of the separation was to prevent listeners from developing relative pitch as a reference (Deutsch et al., 2006). The 36 tones of a given timbre were divided into 3 blocks of 12 tones, with a 5 s interstimulus interval and a 10 s break between the blocks. Notes of a given timbre were always presented together. Six lists of stimulus presentation were created to counterbalance the order of timbre presentation. The participants were randomly assigned to one of the six lists.

2. Participants

The 36 musicians described in Experiment 1 participated in this experiment.

3. Procedure

BLISS was used for stimulus presentation. Participants were tested individually in the same sound-treated booth with
the same pitch range (C3 to B5) and were 500 ms long. As in were similar to the test tones in that they were selected from the same pitch range (C3 to B5) and were 500 ms long. As in the actual experiment, any two consecutive tones were separated by more than an octave. None of the practice tones appeared in the actual experiment. No feedback was given.

Prior to the experiment, 12 practice trials of synthesized oboe tones were given to familiarize the participants with the presentation and response format. The practice tones were similar to the test tones in that they were selected from the same pitch range (C3 to B5) and were 500 ms long. As in the actual experiment, any two consecutive tones were separated by more than an octave. None of the practice tones appeared in the actual experiment. No feedback was given.

B. Results and discussion

Participants’ responses were graded manually. Following previous studies (Lee and Hung, 2008; Lee and Lee, 2010; Lee et al., 2011), four accuracy measures were derived from the response data, including percentages of correct identification allowing zero-, one-, two-, and three-semitone errors. Figure 2 shows the accuracy of identification as a function of timbre and number of semitone errors allowed. The figure shows that none of the musicians possessed absolute pitch according to the 85% criterion defined in Deutsch et al. (2006). Identification remained at chance level even when the scoring criterion was relaxed to allow for semitone errors. The same result was obtained in our earlier study with English-speaking musicians (Lee and Hung, 2008), which is consistent with the low occurrence of absolute pitch in nontone language speakers (Deutsch, 2013). This finding contrasts with studies showing high prevalence of absolute pitch in tone language speakers (Deutsch et al., 2006; Deutsch et al., 2009; Lee and Lee, 2010; Lee et al., 2011), suggesting an association between absolute pitch and tone language experience.

To evaluate the effect of timbre, one-way repeated measures ANOVAs were conducted on the four accuracy measures with timbre (piano, pure tone, and viola) as a within-subject factor and participants as a random factor. When a main effect from the ANOVAs was significant, the Bonferroni post hoc test was used for pairwise means comparisons to keep the familywise type I error rate at 5%. The results confirmed that the musicians’ identification performance was at chance level in all four accuracy measures. In addition, no effect of timbre was found except that when three-semitone errors were allowed, identification accuracy was higher for piano than for pure tone. This result is similar to our earlier study with English-speaking musicians (Lee and Hung, 2008), where timbre had essentially no impact on identification accuracy. This finding, however, contrasts with previous studies showing a timbre effect in individuals with absolute pitch (see Deutsch, 2013, for a summary). It may be that the presence of timbre (e.g., piano) facilitates musical tone identification without a reference pitch only for individuals with absolute pitch.

To test the hypothesis that pitch identification without a reference pitch is associated with the age of onset of musical training, Pearson’s correlation coefficients were derived between the four dependent measures and the age of onset of musical training. Fisher’s $r$ to $z$ transformation was then carried out to evaluate whether the correlation coefficients were significantly different from zero. The coefficients ranged from $-0.3$ to $0.7$. None of them were statistically significant. In other words, there was no evidence for the association between musical tone identification and age of onset of musical training in this sample of musicians. Similar to the English-speaking musicians in Lee and Hung (2008), the average age at which our participants started musical training was close to nine years. Deutsch et al. (2006) reported that none of their English-speaking musicians with a similar age of onset possessed absolute pitch. By contrast, previous studies on individuals with absolute pitch consistently showed a negative correlation between identification accuracy and age of onset of musical training (Deutsch et al., 2006; Deutsch et al., 2009; Lee and Lee, 2010; Lee et al., 2011). Taken together, these findings are consistent with the observation that the prevalence of absolute pitch is negatively associated with the age of onset of musical training.

IV. ASSOCIATIONS BETWEEN LEXICAL AND MUSICAL TONE IDENTIFICATION

Data from the two experiments allowed for an evaluation of the association between pitch height perception in lexical and musical tones. Identifying isolated Taiwanese level tones without external cues for speaker normalization requires comparing a stimulus to a listener’s internal representation of speakers’ vocal ranges. Identifying a musical tone without a reference pitch also requires comparing stimulus input to a listener’s internal scale of pitch without a reference pitch. If these two abilities are associated, there should be a positive correlation between performances in these two tasks.

To evaluate this hypothesis, Pearson’s correlation coefficients were calculated between accuracy measures of Taiwanese tone identification and those of musical tone
identification. The music measures included percentages of correct responses to piano, pure tone, and viola stimuli, as well as mean accuracy averaged across all musical-tone stimuli. The Taiwanese measures included percentages of correct responses to high- and mid-tone stimuli produced by female and male speakers, as well as mean accuracy averaged across all Taiwanese-tone stimuli. Fisher’s r to z transformation was carried out to test whether the correlation coefficients were significantly different from zero. Given the chance-level performance in the music task, no meaningful correlations between the lexical and musical tasks were expected. As expected, the correlation coefficients were generally low and none of them were significantly different from zero.

Although performance in the music task was generally at chance level, Fig. 2 shows a few outliers who attained higher accuracy. It is possible that an association between musical and lexical tone identification could be found in these musicians. Therefore, we examined these outliers separately. Overall, considering the accuracy of responses to piano stimuli with zero semitone errors allowed, there was one musician at 58.3%, one at 22.2%, and the rest were below 20%. We examined the data in three ways. First, since chance performance is 8.3%, we divided all musicians into two groups: above chance (11 musicians) and below/chance (25 musicians). Unpaired t-tests were conducted to compare the performance of the two groups using five dependent measures of lexical tone identification (female/high, female/low, male/high, male/low, and all stimuli combined). None of the tests revealed significant differences between the two groups. Second, we looked at the two musicians who attained 58.3% and 22.2% accuracy in the piano task. If performance in the music task predicts performance in the lexical tone task, these two musicians should rank relatively high in the lexical tone task. However, they ranked 22nd and 27th among the 36 musicians. Third, using only the 11 musicians who achieved above-chance performance in the piano task, we calculated correlations between performance in the piano task and that in the lexical tone tasks (using the same five dependent measures). None of the five correlation coefficients were significantly different from zero.

Taken together, these additional analyses did not provide evidence for the association between musical and lexical tone identification. Since none of the musician participants in this study actually possessed absolute pitch, it is possible that an association could be found in individuals with absolute pitch. However, our previous studies of Mandarin- and Taiwanese-speaking musicians with absolute pitch also failed to show evidence for such an association. The consistent absence of correlation suggests that the ability to identify pitch height for lexical and musical tones may not be related.

V. GENERAL DISCUSSION

Two experiments were conducted in this study to evaluate English-speaking musicians’ and nonmusicians’ use of pitch height to identify Taiwanese level tones and musical tones. The Taiwanese task showed that the musicians had a slight advantage over the nonmusicians. Both the musicians and nonmusicians were able to identify tone height above chance, but only for tones at the extremes of the speakers’ overall vocal range. The music task showed that none of the musicians met the criterion for absolute pitch. Timbre of the musical-tone stimuli did not affect how accurately the tones were identified. Age of onset of musical training did not correlate with performance in the music task either. There was also no evidence for the association between performance in the Taiwanese task and in the music task. Based on the results from this and our previous studies (Lee and Hung, 2008; Lee and Lee, 2010; Lee et al., 2011), we now discuss the implications of these findings for the roles of tone language experience and musical training in lexical and musical pitch perception.

Overall, among the nontone language speakers, the musicians identified lexical tones with higher accuracy than the nonmusicians. However, the advantage diminished as the amount of acoustic information was reduced. Musicians identified intact and silent-center Mandarin tones with higher accuracy than nonmusicians, but there was no difference for onset-only tones (Lee and Hung, 2008). The present study further showed that the musicians’ advantage over nonmusicians in Taiwanese level tone identification was minimal (3% on average). In other words, musicians had a substantial advantage in lexical tone identification only when tonal distinctions involved dynamic F0 contours, and only when the contour information was available (as in the case of intact Mandarin tones) or could be inferred (as in the case of silent-center Mandarin tones, where F0 contour could be inferred from the starting and ending F0). By contrast, when F0 contour information was not available (as in the case of onset-only Mandarin tones and Taiwanese level tones), the musicians’ advantage was either absent (Lee and Hung, 2008) or minimal (the current study). Taken together, these findings suggest that the musicians’ advantage in lexical tone perception is largely due to their ability to track F0 contours more effectively than the nonmusicians. Furthermore, there was no correlation between performance in the music task and that in the lexical tone task, suggesting the ability to identify pitch height in lexical tones is not associated with the ability to identify pitch height in musical tones. Nonetheless, the slight advantage of the musicians in Taiwanese level tone identification suggests that musical training still contributes to pitch height identification. Our speculation is that musicians, like tone language speakers, may be better than nonmusicians at using inherent covariances between F0 and voice quality to infer pitch height. This point will be elaborated shortly.

Since none of the English-speaking musicians in Lee and Hung (2008) and the current study actually possessed absolute pitch, it is possible that an association between pitch height detection in lexical and musical tones could emerge in musicians with absolute pitch. However, the interpretation seems unlikely in light of the data from musicians speaking a tone language. Lee and Lee (2010) found that 72% of the 72 Mandarin-speaking musicians possessed absolute pitch. Lee et al. (2011) found that 65% of the 43 Taiwanese-speaking musicians possessed absolute pitch. That is, both samples included substantial numbers of absolute pitch possessors, as would be predicted by Deutsch’s proposal of the association between absolute pitch and tone.
language. However, no correlations were found between the two tasks in either study, whether all musicians or just those with absolute pitch were included in the correlation analysis. All findings considered, the ability to track F0 contours remains the most likely explanation for the musicians’ advantage in identifying lexical tones.

Turning to the role of tone language experience on pitch height identification, the data from the English listeners in this study showed a distinct pattern from the data from Taiwanese listeners reported in Lee et al. (2011). Whereas the Taiwanese listeners were able to identify all four classes of stimuli (high tones produced by female speakers, HF; mid tones produced by female speakers, MF; high tones produced by male speakers, HM; mid tones produced by male speakers, MM) with accuracy above chance, the English listeners were able to identify only HF and MM stimuli above chance. In addition, for the Taiwanese listeners, the accuracy was comparable among the four classes of stimuli (with the exception of HF stimuli, which were identified better than the other three groups). For the English listeners, by contrast, HF and MM stimuli were identified significantly better than MF and HM stimuli. In other words, the English listeners were able to identify the level tones reliably only when the stimulus tones were at the extremes of the speakers’ overall vocal range. The Taiwanese listeners, by contrast, were able to identify all four groups of stimuli with above-chance accuracy.

While this finding clearly indicates that tone language experience plays a role in how lexically contrastive pitch height is processed, further explanations are needed for why the response patterns are different between the English and Taiwanese listeners. The English listeners’ response pattern seems intuitive. Tones near the extremes of the vocal pitch range were identified more accurately because, as Fig. 1 shows, there is less overlap between the high and mid tones at the extremes (HF and MM stimuli) compared to those in the middle of the vocal range (MF and HM stimuli). It makes sense for the stimuli that are acoustically more distinct to be identified more accurately.

How, then, do we account for the above-chance performance of the Taiwanese listeners in identifying all four classes of stimulus, including those in the middle of the vocal range? Since typical cues for speaker normalization (F0 contour, context, and familiarity with speakers) were removed from the stimuli, the listeners were not likely to rely on cues extrinsic to the stimulus syllables. That means their success was most likely due to utilizing acoustic information within the syllables and mapping that information onto pitch templates stored in long-term memory. The questions, then, are what information other than F0 could be available for tone height identification, what kind of pitch templates are stored in memory, and how is the mapping achieved?

Regarding acoustic cues other than F0, Lee’s (2009) acoustic analysis of onset-only Mandarin tones showed that the high-onset tones (high-level and high-falling) and low-onset tones (mid-rising and mid-dipping) are significantly different in F0, duration, and two voice quality measures (F1 bandwidth and spectral tilt). Correlation analysis further showed that F0 covaried with the voice quality measures and tone classification based on height also correlated with the voice quality measures. The systematic relationship between voice quality and F0 suggests that the covariation could be a source of information used by listeners to infer tone height. If this is true, one would expect similar associations between F0 and voice quality measures in Taiwanese. We are currently investigating this possibility.

As for the nature of pitch templates stored in memory and the mapping from the acoustic signal onto the templates, it has been proposed that pitch class templates can be acquired from exposure to prevalent speaking F0s of a linguistic community (Deutsch et al., 1990; Deutsch, 1991; Deutsch et al., 1999, 2004). Lee (2009) and Lee et al. (2010) further proposed that gender identification contributes to pitch height estimation. Specifically, listeners possess pitch templates based on mean F0s for female and male speakers that the listeners have experienced throughout their lives. Listeners also possess the knowledge of the covariations between F0 and voice quality measures. Upon acoustic input, listeners identify speaker gender based on mean F0. Gender-specific pitch templates stored in memory are then used to evaluate the acoustic input. During the evaluation, knowledge of the F0-voice quality covariations is exploited to calibrate the pitch height estimation. This proposal is supported by the finding that pitch quality measures consistently distinguished female from male Mandarin stimuli in Lee (2009). It is also supported by the finding that both Mandarin and English listeners were able to identify speaker gender reliably from onset-only Mandarin stimuli (Lee et al., 2010). We suspect that the listeners’ success in identifying Taiwanese tone height in the present study involved the same process, i.e., identifying speaker gender as a precursor to estimating tone height. In addition, the Taiwanese listeners’ advantage over the English listeners is most likely due to their relative familiarity with the prevalent speaking F0s and the F0-voice quality relationship in their linguistic community. This proposal could be tested by examining gender identification from the Taiwanese-tone stimuli. Successful gender identification from the same set of stimuli would provide further support for the role of gender identification in pitch height estimation.

An implicit message from the above discussion is that lexical tone perception involves more than F0 processing. Acoustic cues other than F0 could play a role when F0 is not sufficient to allow tone height estimation. It is known that duration and amplitude contours are cues to lexical tone perception, particularly when F0 information is not available (Whalen and Xu, 1992; Liu and Samuel, 2004). These observations indicate that listeners are able to use their knowledge of the internal category structure of lexical tones for identification when acoustic input and listening conditions are less than optimal, such as identifying onset-only Mandarin tones and Taiwanese level tones without cues for speaker normalization.

Could a similar process account for the performance in the music task? That is, could timbre play a role in musical pitch height identification, as voice quality does in lexical pitch height identification? There is ample evidence showing the effect of timbre on musical pitch identification without a reference pitch (see Deutsch, 2013, for a summary). The general finding is that piano stimuli are more effective in
facilitating note naming. This result is consistent with our studies of Mandarin-speaking musicians (Lee and Lee, 2010) and Taiwanese-speaking musicians (Lee et al., 2011). By contrast, our studies of the English-speaking musicians (Lee and Hung, 2008, and the present study) did not show the effect of timbre. Since the majority of the tone-language-speaking musicians possessed absolute pitch and none of the English-speaking musicians possessed absolute pitch, it could be that tone language experience, absolute pitch, or both contributed to the timbre effect. Since the timbre effect has been reported in absolute pitch possessors without tone language experience (Deutsch, 2013), our conjecture is that musicians with absolute pitch might be able to use the covariations between F0 and timbre more effectively to identify musical pitch height. McAdams (2013) noted that the spectral properties of many instrumental sounds covary with pitch. There is also evidence that pitch perception is influenced by timbre perception (see McAdams, 2013, for a review). These findings suggest that both lexical and musical pitch perception may involve the use of covariations between pitch and other acoustic properties. Further research could test that hypothesis to further explore the music-language relationship.

VI. CONCLUSION

Tone languages present a unique opportunity for examining the relationship between music and speech/language because of the lexical use of pitch. This study showed that English-speaking musicians and nonmusicians were able to identify Taiwanese level tones above chance for stimuli at the extremes of the speakers’ vocal range. However, none of the musicians qualified for absolute pitch, and no correlations were found between lexical and musical pitch perception. Results from this series of studies on lexical and musical pitch perception suggest that musicians’ advantage in lexical tone perception is due to their ability to track F0 contour effectively. The findings also suggest that gender identification as well as knowledge of covariations between F0 and voice quality/timbre may contribute to the identification of pitch height in lexical and musical tones.

ACKNOWLEDGMENTS

This research was supported in part by the Provost’s Undergraduate Research Fund from Ohio University and a professional development fund from the Honors Tutorial College at Ohio University.


