Enhanced production and perception of musical pitch in tone language speakers

PETER Q. PFORDRESHER
University at Buffalo, State University of New York, Buffalo, New York

AND

STEVEN BROWN
McMaster University, Hamilton, Ontario, Canada

Individuals differ markedly with respect to how well they can imitate pitch through singing and in their ability to perceive pitch differences. We explored whether the use of pitch in one's native language can account for some of the differences in these abilities. Results from two studies suggest that individuals whose native language is a tone language, in which pitch contributes to word meaning, are better able to imitate (through singing) and perceptually discriminate musical pitch. These findings support the view that language acquisition fine-tunes the processing of critical auditory dimensions in the speech signal and that this fine-tuning can be carried over into nonlinguistic domains.

Pitch plays a prominent role in the structure of both speech and music. For music, pitch conveys information about tonality (e.g., Krumhansl, 1990), harmonic changes (e.g., Holleran, Jones, & Butler, 1995), phrase boundaries (e.g., Deliège, 1987), rhythm and meter (e.g., Jones, 1987), and the likelihood of future events (e.g., Narmour, 1990). For speech, it conveys information about word stress, utterance accent, pragmatic stance, and speaker emotion (Cruttenden, 1997). For the large class of languages known as tone languages, pitch conveys information about lexical meaning as well (Yip, 2002).

Given the vital importance of pitch for human communication, it is intriguing to note the degree to which individuals vary in their ability to produce and perceive pitch. Such individual differences are most apparent in studies of singing. A majority of individuals claim to be unable to imitate melodies by singing (Pfordresher & Brown, 2007) and therefore often describe themselves as "tone deaf" (Cuddy, Balkwill, Peretz, & Holden, 2005; Sloboda, Wise, & Peretz, 2005; Wise & Sloboda, 2008). Although poor singing appears to be less prevalent than self-assessments would indicate, inaccurate singing does affect a sizable minority of the population: 15%–20% of nonmusicians (Dalla Bella, Giguère, & Peretz, 2007; Pfordresher & Brown, 2007). With regard to perception, tasks as straightforward as pitch discrimination show significant individual differences as well (e.g., Amir, Amir, & Kishon-Rabin, 2003), and a small subset of the population (perhaps 5%) appears to be deficient in the discrimination of musical pitch (e.g., Peretz et al., 2002). Whereas musical pitch processing is clearly mediated by multiple cogni-

tive processes (Peretz & Coltheart, 2003), our interest in the present study was to examine one potential mediating variable: the tonal properties of one's native language. The research summarized above did not incorporate linguistic background as a potential factor in individual differences, although there is reason to believe that tone languages may facilitate musical pitch processing.

The world's languages can be classified as either tone or intonation languages on the basis of whether they employ pitch to convey word meanings (Yip, 2002). Tone languages, which represent roughly two thirds of all languages in use, use pitch height and/or pitch changes to convey word meanings, whereas intonation languages do not. A popular example from the Chinese tone language Mandarin is that the syllable /ma/ can be associated with at least four different unrelated meanings on the basis of the tone used. For some tone languages, including many of the Asian tone languages, lexical information is conveyed via contours (i.e., pitch intervals), whereas in others, including many African tone languages, lexical information is conveyed via pitch height rather than contours per se (Yip, 2002). An individual who acquires a tone language must form strong associations between pitches and word meanings both for speech production and speech perception; possibly as a result, adult speakers of tone languages produce highly consistent pitch contours (Deutsch, Henthorn, & Dolson, 2004).

In contrast, intonation languages, such as English, employ pitch to convey information about syllable stress, sentence focus, and speaker attitude, such that mappings between pitch and meaning are much more coarse grained

P. Q. Pfordresher, pqp@buffalo.edu

than is typically the case for tone languages. Pitch is used to convey intonational and pragmatic meanings at the utterance level rather than lexical meanings at the word level (Cruttenden, 1997; Wennerstrom, 2001). Taken to an extreme, speakers of intonation languages may communicate word meanings effectively with minimal or even inappropriate use of pitch contours, as seen, for example, in autism spectrum disorders (McCann & Peppé, 2003).

On the basis of these differences between tone and intonation languages, we hypothesized that individual differences in pitch processing might vary as a function of native language. More specifically, we hypothesized that the stronger requirement for pitch precision in a tone language would carry over to the nonlinguistic context of musical pitch processing, as demonstrated by an enhancement of performance by tone language speakers (when musical background is controlled for) on musical pitch tasks, including production (imitation) and perception (discrimination).

Furthermore, we wanted to examine two different representations of musical pitch: absolute and relative pitch. *Absolute pitch* refers to the categorical representation of pitch that exists independent of any contextual information. Although the ability to label the absolute pitch of a single tone is rare (e.g., Levitin & Rogers, 2005; Takeuchi & Hulse, 1993; Ward, 1999), many individuals are able to reproduce the absolute pitch of popular songs while singing (Levitin, 1994) and can recognize absolute pitch from television theme songs (Schellenberg & Trehub, 2003). In the present article, we use the term *absolute pitch* to refer to the ability to imitate or discriminate individual notes on the basis of their pitch class, irrespective of their relationship with surrounding notes. By contrast, *relative pitch* refers to relationships among pitch classes in a melody, irrespective of each individual pitch's class. For instance, the melodic intervals C4–G4 (pitches C and G in octave 4) and E4–B4 both form seven-semitone intervals, even though each pair comprises different pitch classes. The ability to represent relative pitch, which is thought to be widespread in humans, allows listeners to hear transposed melodies as similar, even though each absolute pitch in the melody differs across transpositions. In this article, we use the term *relative pitch* to refer to the processing of intervals—changes in pitch from one note to the next—independent of each individual note's pitch.

Some research suggests that tone language speakers show pitch-processing advantages in nonlinguistic domains. For example, Chinese speakers, in addition to showing categorical perception of linguistic tone, demonstrate categorical perception of nonspeech tone analogues, whereas English speakers do not (Xu, Gandour, & Francis, 2006). Moreover, Chinese music conservatory students show much higher occurrences of absolute pitch labeling abilities than do American music conservatory students (Deutsch, Henthorn, Marvin, & Xu, 2006). One theory proposes that a major neural hub for auditory–motor integration (located at the junction of the temporal and parietal lobes) is shared between music and speech, suggesting that production–perception links acquired through language may in fact transfer to music (Hickok, Buchsbaum, Humphries, & Muftuler, 2003). Certain aspects of phonological production in speech may therefore be guided by general audio–motor mechanisms rather than by speech-specific mechanisms.

However, other results fail to support the idea that linguistic background influences the processing of musical pitch. Recent neuroimaging research suggests that tone language speakers use separate neural networks for the perception of linguistic pitch as opposed to pitch in other contexts (like music). Specifically, the discrimination of lexical tones in a linguistic context (as opposed to a nonlinguistic context) increases activations in the left inferior frontal gyrus (Broca's area) for tone language speakers but not for individuals unfamiliar with the language (including intonation language speakers; Gandour et al. 2000; Gandour, Wong, & Hutchins, 1998; Wong, Parsons, Martinez, & Diehl, 2004). These results follow from a long-standing idea that speech perception relies on speech-specific neural mechanisms (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985; Peretz & Zatorre, 2005; Remez, Rubin, Berns, Pardo, & Lang, 1994; Trout, 2001; but see Galantucci, Fowler, & Turvey, 2006, for a critique of this perspective). Similarly, a recent neuropsychological model of music and language includes the claim that the processing of pitch may be performed by domain-specific and independent modules, depending on whether pitch appears in a linguistic or a musical context (Peretz & Coltheart, 2003).

Given such different findings, it is not surprising to note that most current theoretical accounts strike a compromise between full integration and full independence of music and language. For instance, Patel (2008) suggests that processing resources are shared between music and language, whereas the representations served by those resources may differ across domains. Similarly, the modular approach proposed by Peretz and Coltheart (2003), although based primarily on the assumption of separation between music and language, suggests that the representation of melodic contour (i.e., direction of pitch change) may in some cases be shared between music and language. In support of this view, research by Patel, Wong, Foxton, Lochy, and Peretz (2008; cf. Patel, 2003) suggests that deficits in musical pitch processing may be linked to an inability to perceive the direction of pitch change and that this inability may extend to pitch contours in speech.

We adopted a new strategy to address whether tone language acquisition facilitates pitch processing in music. First, unlike in much previous work that has addressed only perception, we included pitch production (vocal imitation) and pitch perception (discrimination) tasks, two skills critical for communication. Second, we analyzed both absolute- and relative-pitch processing. For both production and perception, this analysis involved comparing participants' accuracy in processing single pitches (absolute pitch) versus accuracy in processing relationships between pitches (relative pitch).

STUDY 1

In Study 1, we compared the music perception and production task performance of individuals whose first lan-
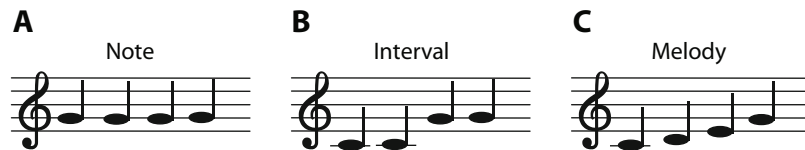
Figure 1. Examples of sequence types used for the vocal imitation tasks, shown in music notation.

guage was an Asian tone language (including Mandarin, Cantonese, and Vietnamese) with that of native speakers of English. In general, we expected to find enhanced pitch processing in tone language speakers compared with intonation language speakers. Some past research, described above, has suggested that the advantage should be greater for relative pitch (e.g., Peretz & Coltheart, 2003), although others have observed facilitating effects for absolute pitch (Deutsch et al., 2006).

Method

Participants

All participants were undergraduate students from the University of Texas at San Antonio (where P.Q.P. was a faculty member). The sample of tone language speakers in this study included 12 students with little or no musical training who participated in exchange for either course credit (in introductory psychology) or payment.[1] Four participants reported modest amounts of musical training ($M = 2.75$ years), none of which involved singing or playing an instrument that required fine-tuning (e.g., a violin). All participants spoke a tone language from East Asia (Vietnamese [$n = 6$], Mandarin [$n = 4$], Cantonese [$n = 2$]) as their first language and were fluent in English as a second language. Their mean age was 19 years (range = 17–30 years). Half of the participants were female, and all but one were right-handed.

Each individual in the sample of tone language speakers was matched with a participant from a larger sample ($n = 40$) of native English speakers who had performed the same tasks in a previous study (Pfordresher & Brown, 2007). Matches were based on gender, age (mean age of comparison group = 21 years, range = 18–39 years), and self-reports of musical skill (excluding singing). One of the participants from the comparison group reported modest musical training (3 years). Overall, the groups did not differ significantly with respect to musical training (mode = 0 years of training for both groups) or self-reports of musical skill. Although participants in each group constituted rough matches to individuals in the other, we adopted a conservative stance in statistical tests and considered the groups to be independent.

Materials and Apparatus

Production trials. Participants vocally imitated four-note sequences (termed *target sequences*). Sequences for male participants were sampled from the pitches C3, D3, E3, F3, and G3 (fundamental frequency, $f0 = 131, 147, 165, 175,$ and 196 Hz, respectively), and sequences for female participants comprised the equivalent pitches one octave higher. Target sequences consisted of a synthesized voice (Vocaloid, Zero-G Limited, Okehampton, England) singing the syllable *da* and were presented over Aiwa HP-X222 headphones at a comfortable listening level. Participants' performances were recorded as .wav files through a Shure SM48 microphone.

Three types of target sequences were created to form three levels of sequence complexity. *Note sequences* ($n = 5$ sequences, one for each pitch), the simplest level, consisted of a pitch repeated four times, as shown in Figure 1A. *Interval sequences*, as shown in Figure 1B, comprised sequences with a single pitch change (i.e., an interval) between Notes 2 and 3. Eight such sequences were generated,

beginning on either C or G and changing to one of the four remaining diatonic pitches on Notes 3 and 4. *Melody sequences*, as shown in Figure 1C, comprised four unique pitches and began on C (four sequences) or G (four sequences). Different melody sequences were formed by varying melodic contour. Two sequences (one ascending, one descending) had no contour changes, like the sequence shown in Figure 1C. Two others had two changes in melodic contour (e.g., C4 E4 D4 G4), and the final four had a single contour change after the second (two sequences) or third (two sequences) note.

Perception trials. Stimuli for the perceptual tasks were created using MIDILAB 5.0 software (Todd, Boltz, & Jones, 1989). Trials involved either note discrimination (comparing pairs of single pitches) or interval discrimination (comparing pairs of two-note intervals). We use the term *note discrimination* rather than *pitch discrimination* in order to draw parallels between these tasks and analyses of production tasks (described below). During note discrimination trials, participants heard two 1-sec sine-wave tones (notes) separated by a 1-sec pause (see Figure 2A). Sine tones were employed because, when pitch discrimination is done with complex tones, it is unclear which frequencies are used to make the discrimination. The first note was always C5 (524 Hz), and the second note could be either the same as (50% of trials) or different from (50% of trials, 25% higher and 25% lower) the first. Pitch changes were calibrated in cents (100 cents = 1 semitone) and included 25-, 50-, 100-, 200-, 400-, 600-, and 800-cent increases or decreases (corresponding, respectively, to differences in frequency of 8, 15, 30, 61, 122, 183, and 250 Hz, averaged across ascending/descending changes).

For interval discrimination trials (Figure 2B), participants heard pairs of two-note melodic intervals (four notes total). The standard interval comprised the pitches C5 and G5 presented in immediate succession. Following a 1-sec pause, the second interval began on F♯5. *No change* trials occurred when the final pitch was C♯6, in which
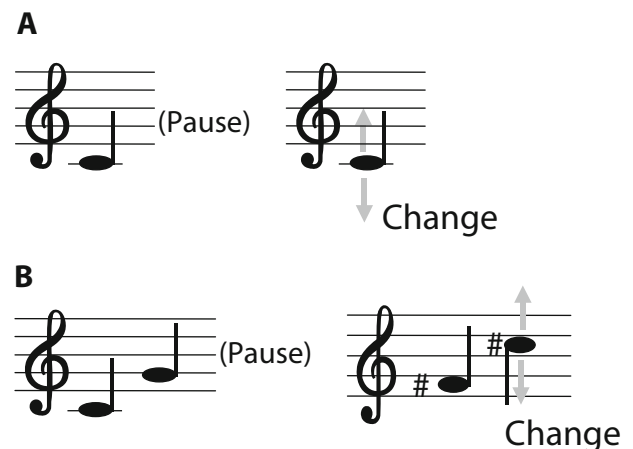

Figure 2. Examples of stimuli used for perceptual discrimination tasks, shown in music notation. (A) Note discrimination. (B) Interval discrimination. Both panels illustrate pitches in *no change* conditions, with locations and directions of changes indicated by arrows.

case both pitch pairs created a 700-cent change. Thus, participants had to use relative pitch in order to judge sameness. This is a very difficult task when isolated intervals are used (Burns & Ward, 1978), even though it can be fairly easy to categorize intervals in familiar melodies (Smith, Kemler Nelson, Grohskopf, & Appleton, 1994). *Change* trials were created by altering the final pitch of the second interval either up or down in pitch relative to C♯6. Modifications were 25, 50, 100, 200, or 400 cents (corresponding, respectively, to differences in frequency of 16, 32, 65, 129, or 260 Hz, averaged across ascending/descending).

Procedure

The study began with production trials, which participants completed in a single-walled sound-attenuated booth (WhisperRoom, Inc., Morristown, TN). Participants stood and were instructed to use their abdominal muscles to support respiration while singing. Participants warmed up by singing "Happy Birthday" and by vocalizing a comfortable pitch (their *comfort pitch*). For production trials, a metronome sounded in order to establish the rate of singing at 1 syllable/sec. During each trial, participants listened to the synthesized voice and then imitated that sequence using the syllable *da*, starting on the metronome beat following a response cue (a bell). This sequence was repeated twice per trial. Production trials were grouped into three blocks by sequence type (note, interval, melody) with different groups of participants experiencing either an ascending order (simple to complex) or the reverse order; exemplars of each sequence type were ordered randomly within each block. After production trials, participants completed questionnaires regarding musical background, linguistic background, hearing abilities, and beliefs about their musical ability.

Perception trials were completed next, with note discrimination trials preceding interval discrimination trials. Before each set of trials (note and interval), participants were given instructions, including extreme examples of *change* and *no change* trials. Instructions for interval discrimination highlighted the need to focus on the difference between pitches rather than on the pitch of any individual note. Participants responded "yes" after each trial if they perceived a change (of pitch or interval) relative to the standard and otherwise responded "no." They did so by pressing one of two buttons on a custom-made keypad. Different change amounts and change directions were randomly ordered within note and interval discrimination blocks; each participant experienced one of two random orders.

Data analysis for production. Mean f0s were calculated through autocorrelation using TF32 software (Milenkovic, 2001). Artifacts (e.g., octave errors) were adjusted manually by altering pitch settings. Produced and target pitch values reported henceforth were based on the mean f0 across each produced syllable in the sequence. (Past analyses of similar data suggest negligible influences of removing the initial consonant or of using median rather than mean f0; Pfordresher & Brown, 2007.) Pitches for the produced and target sequences were converted from hertz into cents (100 cents = 1 semitone) relative to the lowest possible pitch in the target sequences (C3 for males and C4 for females) to reflect pitch accuracy.

Two error measurements were computed for performance on each of the four-note sequence types. *Note errors* were determined by calculating the difference between the mean produced f0 and the target f0 for each produced pitch. When the sign is kept, positive note errors indicate sharp performance and negative errors indicate flat performance. Note errors function as a measure of how well participants imitate absolute pitch information, in that the errors are based on the reproduction of single pitch events (i.e., "notes"). *Interval errors*, our second error measure, were computed in two steps. First, the f0 difference between successive pitches was computed for each target interval and produced interval. Next, the size of each target interval was subtracted from the size of each produced interval. When the sign is kept, positive errors indicate "overshooting" of the target intervals and negative errors indicate "undershooting." Interval errors function as a measure of how well participants imitate *relative pitch*, in that errors are independent of the accuracy with

which they match the absolute pitches within the intervals. Hence, if a participant were to transpose all the intervals by a fixed amount, their interval error would be zero; their note error would reflect the magnitude of the transposition.

In order to treat overshoots and undershoots as comparable errors, we focused on absolute errors (except where specified). Specifically, we used the absolute value of the produced error for each note or interval and then averaged them to generate absolute note error or absolute interval error, respectively. We should point out that this type of measurement, unlike signed error measures, is influenced by both accuracy (average proximity to target) and precision (variability of performance) and is thus typically larger than signed error scores (Schmidt & Lee, 1999).

Results

Comparisons were carried out between results from tone language speakers and the matched sample of intonation language speakers. There was no indication of differences among speakers of tone languages (Mandarin, Cantonese, or Vietnamese). A comparison of tone language speakers with all the participants from the earlier study (Pfordresher & Brown, 2007) led to the same conclusions. Groups did not differ significantly with respect to their mean produced comfort pitch. The mean absolute difference between singers' comfort pitches and the mean pitch of the target stimuli did not differ significantly across language groups (tone language group, $M = 390$ cents, $SE = 103$; intonation language group, $M = 596$, $SE = 105$). We also assessed accuracy in producing intervals for "Happy Birthday" by comparing each produced interval with its corresponding interval in the original song (see the Data Analysis for Production section). Note that this analysis disregards the key in which the song is sung. Groups did not differ in their interval error for this familiar song (absolute error for tone language speakers, $M = 108$ cents, $SE = 9.8$; for intonation language speakers, $M = 129$, $SE = 18.1$).[2] Thus, it is unlikely that any group differences observed with our imitation tasks would reflect differences in comfortable vocal range or differences in relative-pitch abilities when performing nonimitative vocal tasks.[3]

As mentioned in the introduction, considerable individual differences exist in musical tasks, particularly for vocal production. Thus, many of our analyses revealed participants who were statistical outliers, defined as having scores more than 2 SDs away from the grand mean (across both language groups). The role of outliers is of interest in this research area, given that poor-pitch singing, or "tone deafness," is a way of characterizing outliers on certain tasks (cf. Ayotte, Peretz, & Hyde, 2002; Dalla Bella et al., 2007; Pfordresher & Brown, 2007; Welch, 1979). Given the importance of outliers, we focus on group analyses for the entire sample, but we also present individual data for each task and, where relevant, present analyses of data after outliers are removed.

Production Tasks

Note errors (absolute pitch). We first consider the accuracy with which participants imitated the pitch of individual notes while imitating four-note sequences, using absolute note errors (described above) as our measurement. We ran a mixed-model ANOVA with group (i.e.,
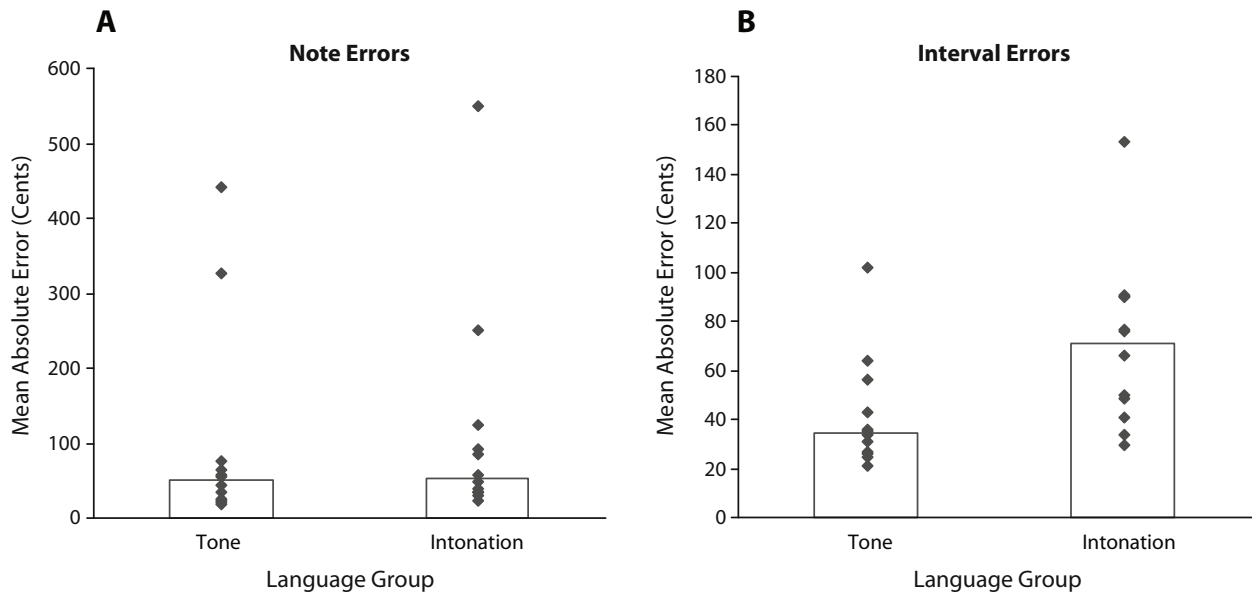
A

B

Figure 3: Absolute note errors (A) and interval errors (B) for individuals (diamonds) and group medians (bars) in Study 1.

language, tone, intonation) as a between-subjects factor and sequence type (note, interval, melody) and serial position of the note (1–4) as repeated measures factors. Our summaries focus on the main effect of group and its interactions with other factors. The ANOVA on absolute note error yielded no main effect of group ($F < 1, p > .10$) and no interactions with this factor. There was a slight advantage, however, for tone language speakers, whose mean absolute errors were lower ($M = 100.08$, $SE = 9.05$) than those for the matched intonation language speakers ($M = 114.95$, $SE = 7.72$).

In order to explore the data further, we assessed individual differences in note errors across the groups. Plots of individual as well as group medians are shown in Figure 3A (medians are used for bars due to the influence of outliers, discussed below), averaged across all sequence types and serial positions. As can be seen, a minority of participants generated errors that were much greater than the median; 1 participant from each group was defined as a statistical outlier, with errors scores more than 2 *SD*s greater than the mean. Nevertheless, the difference across groups was not significant when assessed using a nonparametric Mann–Whitney *U* test ($p > .10$), which minimizes the influence of outliers, and removing these participants did not lead to a significant ANOVA result.

Interval errors (relative pitch). We next turn to the accuracy with which speakers of tone languages and intonation languages imitate musical intervals while imitating four-note sequences, using the absolute interval error (see the Data Analysis for Production section). This measure revealed a significant advantage for tone language speakers. An ANOVA run with the same design as that used for note errors revealed a main effect of group [$F(1,22) = 4.91, p < .05$] and a group × sequence type interaction [$F(2,44) = 8.03, p < .01$]. The significant interaction is shown in Figure 4 (means and medians are highly similar

for interval errors, in contrast to note errors). For this analysis, the *note sequence* type addresses how well participants can maintain a regular pitch across repeated productions. Although both groups generated larger errors for more complex sequence types, this increase was steeper among speakers of English. Post hoc tests (Tukey's HSD, $\alpha = .05$) suggested that the interaction reflects a higher error rate for intonation language speakers' reproductions of melody sequences, which differed from all other conditions. No other pairs differed according to the post hoc test.

As seen in the analysis of absolute note error, certain participants deviated noticeably from the rest of the group with regard to absolute interval error, and 1 intonation language speaker was a statistical outlier (Figure 3B). We therefore confirmed the group difference with a Mann–
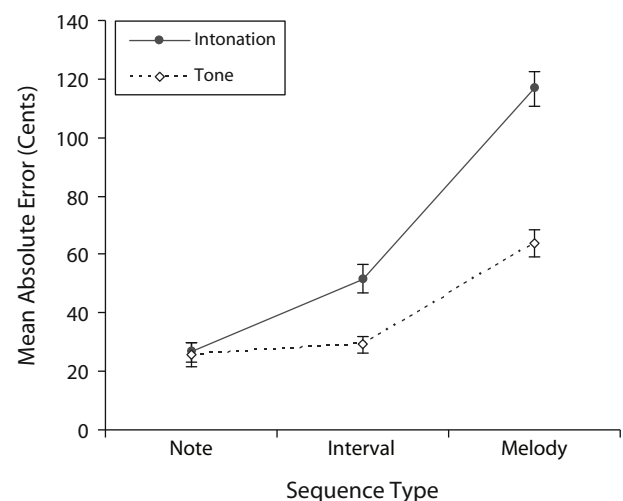

Figure 4. Mean absolute interval error in Study 1, as a function of sequence type. Error bars represent ±1 *SE*.

Whitney U test, which was also significant ($U = 113, p < .01$). As with note errors, removal of the outlier did not change the significance of ANOVA results. It is also worth noting that the intonation language speaker who was a statistical outlier with respect to interval errors (152 cents) also produced the largest note errors (549 cents). By contrast, the tone language speaker who generated the highest absolute interval error (102 cents) was not deficient with respect to note reproduction (53 cents error).

Perceptual Tasks

Note discrimination (absolute pitch). We now turn to perceptual discrimination of individual notes. Discrimination abilities were assessed in a way that separates response bias from accuracy. Due to the presence of participants who made very few errors, we used a difference metric in lieu of d'—namely, hit (H) minus false alarm (FA) rate. This metric has been used in related research on congenital amusia (e.g., Ayotte et al., 2002). Individual means and group medians are shown in Figure 5A. These data were analyzed using a one-tailed t test, averaged across change amounts. No tone language advantage was evident. Subsequent analyses revealed no effects of change magnitude or change direction (upward versus downward). The lack of a group difference is unlikely to have resulted from a ceiling effect; a t test contrasting all scores (both groups combined) with perfect performance ($H - FA = 1$) was significant [$t(23) = -5.87, p < .01$]. Two participants (those generating the worst performance in each group) were identified as statistical outliers; their removal did not influence the effect of group on note discrimination. Of these outliers, 1 (the intonation language speaker) was also an outlier on the absolute note error measure.

Interval discrimination (relative pitch). We analyzed accuracy for interval discrimination tasks just as we

did for note discrimination. Interval discrimination is difficult for nonmusicians (Burns & Ward, 1978; cf. Smith, 1997); nevertheless, performance across groups was significantly better than chance [$H - FA = 0$; $t(23) = 8.18, p < .01$]. In contrast to the result for note discrimination, tone language speakers were significantly better at discriminating intervals than were intonation language speakers, as revealed by an analysis of $H - FA$ difference scores [$t(22) = 2.99, p < .01$] (see Figure 5B). The difference was also significant according to a nonparametric Mann–Whitney U test ($U = 116, p < .01$). As with note discrimination, subsequent analyses revealed no influence of change magnitude or change direction (second interval larger or smaller than the first interval). No participants were identified as statistical outliers, which is due mostly to the fact that performance in general was more variable for interval discrimination than for note discrimination.

Perception and production combined. The fact that tone language speakers demonstrated an advantage for both the imitation (production) and discrimination (perception) of intervals suggests that the acquisition of a tone language jointly facilitates production and perception, at least for relative-pitch information. We addressed whether this perception–production link existed across individuals by correlating scores across the perception and production tasks. Absolute interval errors during production (averaged across trials for each individual) were significantly correlated with $H - FA$ difference scores for interval discrimination when both samples were pooled together [$r(22) = -.57, p < .05$]. Within groups, this relationship was significant only for intonation language speakers [$r(10) = -.60$]. Not surprisingly, correlations between absolute note errors for production and $H - FA$ difference scores for note discrimination were nonsignificant when they were pooled across groups, as they were for tone lan-
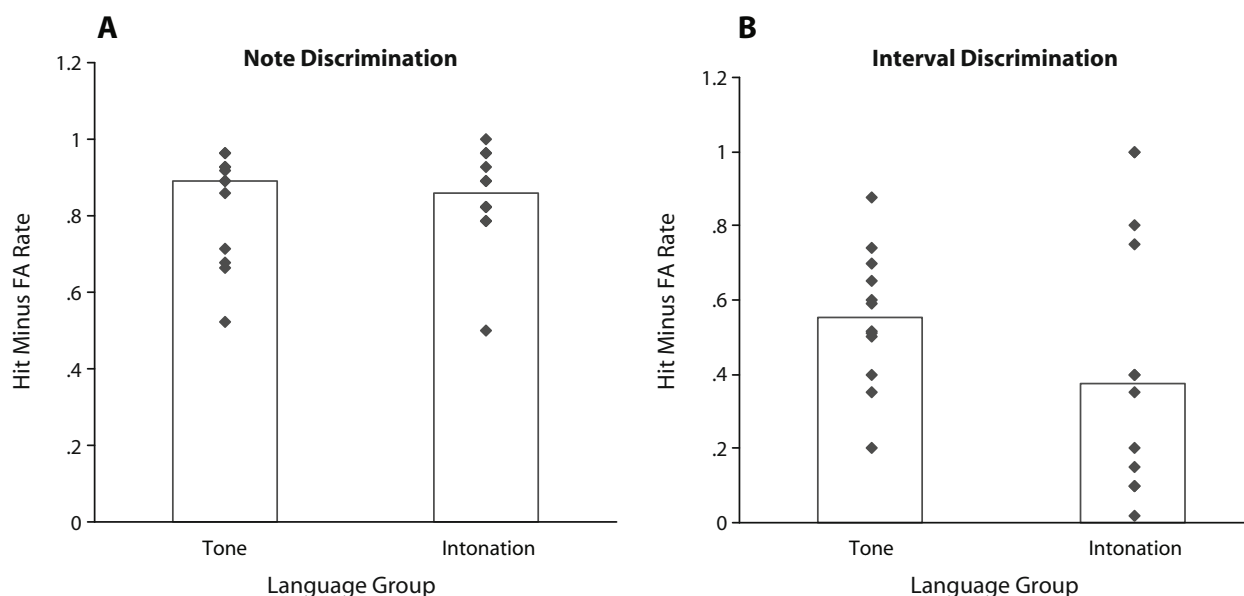

Figure 5. Difference scores between hit rates and false alarm (FA) rates for note discrimination (A) and interval discrimination (B) in Study 1. Diamonds plot individual performance and bars represent group medians.

guage speakers. The correlation was significant for into-nation language speakers [$r(10) = -.65$]. However, the result was attributable to the statistical outlier identified above (who was an outlier on both measures) and became nonsignificant when this individual was removed.

Discussion

Results from Study 1 demonstrated a clear advantage for native speakers of tone languages relative to individuals who speak only intonation languages (here, American English) in both the production and perception of musical intervals, whereas they showed only a slight and unreliable advantage when it came to the production of individual pitches and no advantage in discriminating individual pitches.

Two limitations of Study 1 led us to attempt to replicate the results in a follow-up study. Both relate to the representativeness of our tone language sample. First, there was a lack of control in the selection of participants. We did not collect data on how long participants spoke their first language before speaking English, nor did we ask how long they had remained in their native country before coming to the United States. Second, we did not obtain any indication of whether the advantage among tone language speakers could be linked to overall cognitive functioning. This is important insofar as a link between musical abilities and general intelligence, although widely debated, may exist (e.g., Schellenberg, 2003; Schellenberg & Peretz, 2007).

STUDY 2

Study 2 was like Study 1 in several respects but kept tighter control over the tone language sample. All participants in Study 2 were speakers of Mandarin, had lived in China for at least 10 years before coming to the United States, and had not learned English during their first 6 years. In addition, by conducting the session for this group in Mandarin, we confirmed that Mandarin participants could still speak Mandarin fluently and comfortably. Study 2 also incorporated a broader range of warm-up vocalization tasks as well as a test of general cognitive ability to address whether any tone language advantage is attributable to nonimitative behaviors. Predictions for Study 2 were identical to those of Study 1.

Method

Participants. A new sample of 22 participants was recruited from the student community at the University at Buffalo. Eleven of the participants were native speakers of Mandarin Chinese, 4 of whom were recruited from the Introduction to Psychology subject pool; the remaining were recruited for pay.[4] (The original Mandarin sample was $n = 12$, but 1 participant with singing experience was dropped.) Mandarin speakers lived in China for 22 years on average before coming to the United States (range, 18–27 years) and began learning English at age 19 on average (range, 7–28 years). They reported being more comfortable speaking Mandarin than English at the time. Of the 11 intonation language speakers, 9 were monolingual native English speakers, recruited from the Introduction to Psychology subject pool. The remaining 2 were bilingual intonation language speakers who were recruited for pay. One was a native speaker of English (who also spoke Japanese) and one a native speaker of Portuguese (an additional native speaker of German was initially recruited but was dropped due to extensive specialized training in auditory perception tasks).

Tone and intonation language speakers were randomly sampled and were not explicitly matched, because we did not want to draw on the same previous data set that we used for intonation speakers in Study 1. Even so, the samples were similar in important respects. As in Study 1, both groups represented similarly low years of musical experience ($M = 2.44$ and 3.44 years for tone and intonation language speakers, respectively; mode = 0 for both) and formal musical training ($M = 2.00$ and 1.44 years for tone and intonation language speakers, respectively; mode = 0 for both), and did not differ significantly from one another on either measure ($p > .10$ for each t test). Groups included similar proportions of male and female participants ($n = 6$ female tone language speakers, $n = 5$ female intonation language speakers, $p > .10$). The sample of tone language speakers was older than the intonation language speakers ($M = 28$ years for tone language speakers; $M = 20$ years for intonation language speakers), but this difference was not significant ($p = .09$).

Apparatus, Materials, and Conditions. Trials in Study 2, unlike in Study 1, were administered through custom-made MATLAB scripts (The MathWorks, Natick, MA) on a custom-made personal computer (JEM Computer Marketing, Buffalo, NY). Participants listened to stimuli during imitation and perceptual-discrimination trials over Sennheiser HD 280 Pro headphones. Imitations were recorded via a Shure PG58 vocal microphone, amplified by a Lexicon Omega Studio preamplifier. To extract pitch contours for data analyses, we used Praat (Boersma & Weenink, 2008) with the same pitch-tracking (autocorrelation) technique as we used in Study 1.

The synthesized-voice sequences used for imitation trials were identical to those we used in Study 1. However, trials did not include a metronome and participants were instructed to respond as soon as possible after a white noise burst that functioned as the response cue. Only interval and melody sequence types were used in Study 2, given that no difference between groups was found for note sequence types in Study 1. Stimuli for note and interval discrimination tasks were sine tones generated by MATLAB via the computer's sound card.

Following perception and production trials, participants completed a questionnaire regarding their musical background (as in Study 1) and a new questionnaire regarding linguistic background that included questions about the years that they lived in China before coming to the United States and the age at which they began to learn English. Participants also completed the Wonderlic Personnel Inventory (Wonderlic, Inc., Libertyville, IL), a test of general intelligence that has yielded strong correlations with the Wechsler Adult Intelligence Survey ($r = .90$; Hawkins, Faraone, Pepple, Seidman, & Tsuang, 1990) but takes less time (12 min) to administer. In order to accommodate differences in the rate at which the test was taken (the test is in English), we counted only responses on the first page (27 items).

Procedure. Participants first completed a set of vocal warm-up tasks, followed by imitation tasks and perceptual discrimination tasks (note then interval stimuli), followed by completion of questionnaires and the Wonderlic Personnel Inventory. Trials for the tone language participants were administered in Mandarin by a native speaker. This functioned as a way of reinforcing the use of linguistic tone during the session and as a way of ensuring that participants were still comfortable using their native language. Similarly, for tone language speakers, "Happy Birthday" was performed in Mandarin (a familiar task for all participants). Participants completed a broader array of warm-up trials in Study 2 than in Study 1, including extemporaneous speech (listing what one had for dinner), a reading passage (the "Rainbow Passage" for English speakers and a Chinese poem for Mandarin speakers), and vocal sweeps (i.e., continuous changes from the lowest to the highest pitch that one can comfortably produce). These additional tasks were used to better estimate vocal range. Unlike Experiment 1, different sequence types for

production trials were presented in random order rather than being blocked by complexity.

Results

Analyses of warm-up tasks were used to determine whether participants were similar in terms of overall vocal range and the fit of their vocal range to the target melodies. With respect to absolute differences between participants' comfort pitch and the mean pitch of target melodies, we found no difference, just as in Study 1 (absolute differences for Mandarin tone language speakers, $M = 505$ cents, $SE = 68$; for intonation language speakers, $M = 479$, $SE = 107$). We also did not find differences in measures of vocal range, including vocal sweeps, and in the vocal range exhibited while reading and speaking extemporaneously. Likewise, correlations between measures from warm-up trials and measures of accuracy in imitation tasks were nonsignificant. As in Study 1, language groups did not differ in the accuracy with which they produced pitch intervals while singing "Happy Birthday" (absolute interval error for "Happy Birthday" among tone language speakers, $M = 97$ cents, $SE = 18.9$; for intonation language speakers, $M = 111$, $SE = 11.3$).

We also analyzed the first 27 items from the Wonderlic inventory (see the Materials section) to address possible differences between language groups with respect to general cognitive ability. All participants completed these items, but only 1 participant got every item correct (performance of each group differed significantly from perfect performance, $p < .01$ for each). The groups did not differ with respect to their performance on these items (M tone = 20.1 items, M intonation = 20.5 items, $p > .10$).

Production Tasks

Note errors (absolute pitch). We used the same measures of performance in Study 2 as in Study 1. We first consider absolute note error (analyzed as in Study 1); Figure 6A shows individual means and the group medians. The ANOVA on all participants failed to yield a main effect of group or a group × sequence type interaction ($p > .10$ for each). However, 1 participant from each group was a statistical outlier (i.e., their error scores were more than 2 SDs higher than the overall mean), and a Mann–Whitney U test on the data for all participants (which reduces the influence of outliers) revealed a significant effect based on medians (tone = 47.50, intonation = 95.01, $U = 93$, $p < .05$). Both bilingual participants from the intonation language group had mean absolute note errors that were higher than the median for the tone language group (M error = 333.74 and 106.76 for those individuals).

Given the implication of the nonparametric test reported above, we reran the ANOVA after removing the two outliers. These participants (1 from each group) are visually apparent in Figure 6A. Removing these participants led to a significant effect of group [$F(1,18) = 7.36$, $p < .05$] and a group × sequence type interaction [$F(1,18) = 5.05$, $p < .05$], as shown in Figure 7A. The main effect of group was confirmed by a Mann–Whitney U test when outliers were removed ($U = 83$, $p < .01$). Post hoc tests of the interaction (Tukey's HSD, $\alpha = .05$) demonstrated that errors among intonation language speakers who reproduced melody sequence types were higher than all other means, which did not differ from each other.

Interval error (relative pitch). Next we consider accuracy in interval imitation. Figure 6B shows individual and group medians for absolute interval errors. An ANOVA on interval errors revealed a main effect of group [$F(1,20) = 5.75$, $p < .05$] but no group × sequence type interaction. This effect was confirmed by a Mann–Whitney U test ($U = 95$, $p < .05$). Both bilingual participants from the intonation language group had mean absolute interval errors
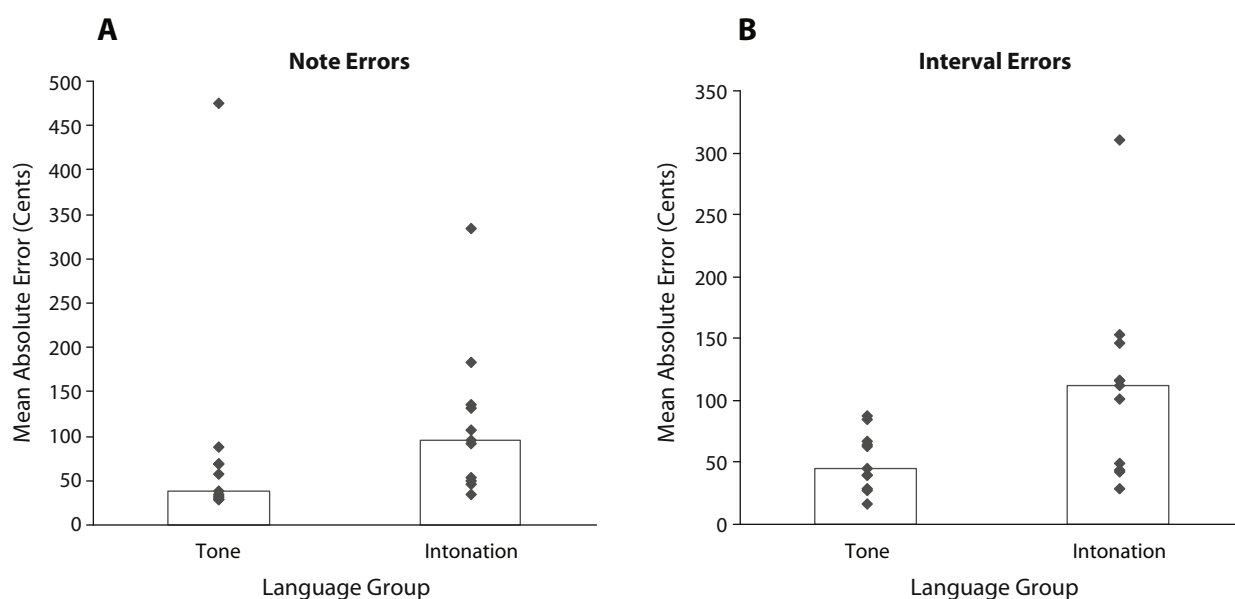

Figure 6. Absolute note errors (A) and interval errors (B) for individuals (diamonds) and group medians (bars) in Study 2.
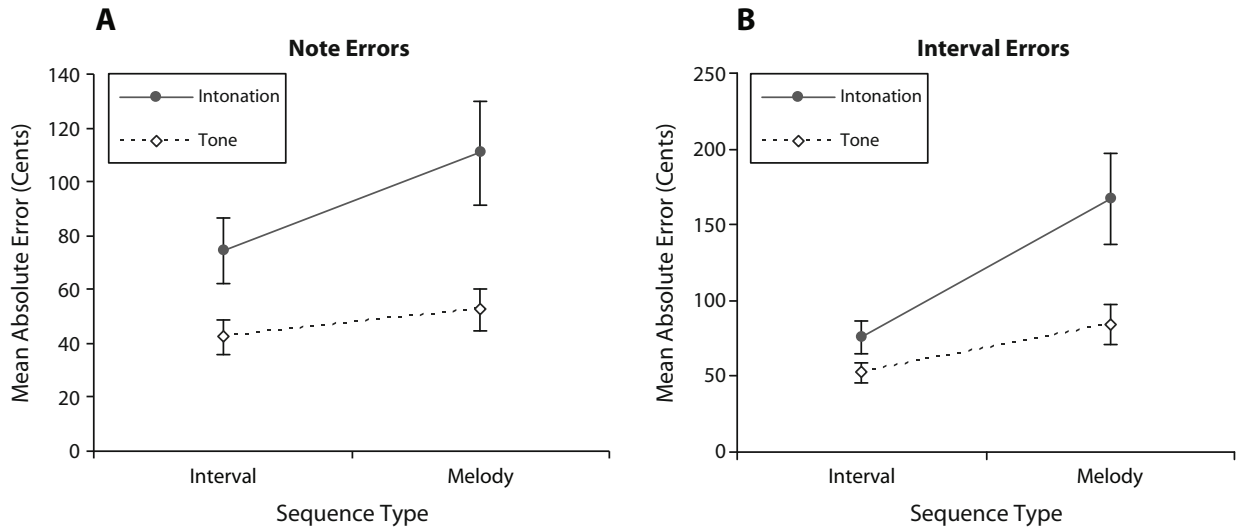
Figure 7. Mean absolute error in production for imitating notes (A) and melodic intervals (B) as a function of language group and sequence type. Error bars represent ±1 *SE*. Each plot represents means after statistical outliers were removed (2 outliers in Plot A, 1 in Plot B).

that were higher than the median for the tone language group ($M = 310.6$, 111.7 for intonation language speakers; median $= 44.93$ for tone language group).

As is apparent in the individual scores of Figure 6B, 1 participant (an intonation speaker) was an outlier. Removal of this participant preserved the main effect of group in the ANOVA [$F(1,19) = 6.43, p < .05$] and also revealed a significant group $\times$ sequence type interaction [$F(1,19) = 6.27, p < .05$], as shown in Figure 7B. Post hoc tests on the interaction revealed a pattern of results like that found for note errors (and also for interval errors in Study 1): Mean absolute errors among intonation lan-

guage speakers reproducing melodies were higher than all other means. The main effect of group after removing the outlier was confirmed by an analogous Mann–Whitney U test ($U = 84, p < .05$).

Perceptual Tasks

Note discrimination (absolute pitch). We now turn to performance on the perceptual discrimination tasks, again analyzed by examining differences between H and FA rates. Means for participants and group medians are shown in Figure 8A. As in Study 1, groups did not differ significantly according to this metric. The lack of a group
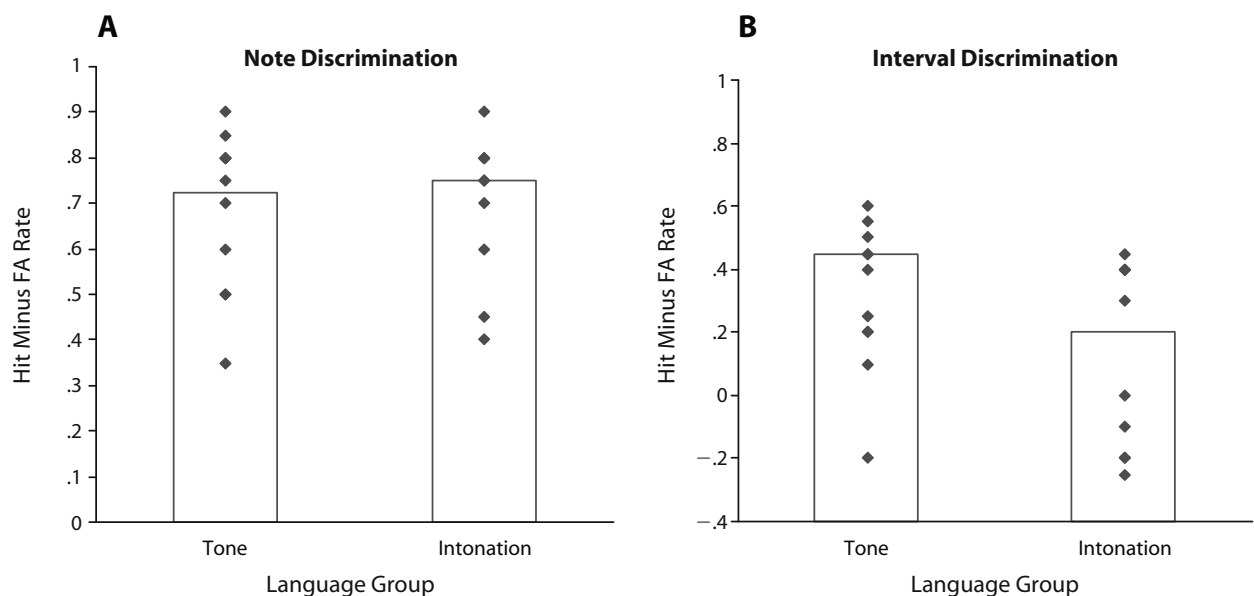

Figure 8. Difference scores subtracting false alarm (FA) rates from hit rates for (A) note discrimination and (B) interval discrimination in Study 2. Diamonds plot individual performance and bars represent group medians.

difference is unlikely to have resulted from a ceiling effect; a t test contrasting all scores (both groups combined) with perfect performance (H − FA = 1) was significant [$t(21) = -8.92, p < .01$]. One tone language speaker (the worst performer in that group) was a statistical outlier; removal of this participant did not alter the results. This participant was not identified as an outlier according to the analogous production measure (note error).

Interval discrimination (relative pitch). As in Study 1, tone language speakers showed an advantage at discriminating intervals, relative to intonation language speakers. This result is shown in Figure 8B. The difference between language groups was significant according to a one-tailed t test [$t(20) = 1.92, p < .05$] and a Mann–Whitney U test ($U = 90, p < .05$). Subsequent analyses revealed no influence of change magnitude or change direction (second interval larger or smaller than the first interval) on the result. As in Study 1, performance across groups was significantly greater than chance [H − FA = 0; $t(21) = 3.94, p < .01$]. Both bilingual participants from the intonation language group had difference scores that were lower than the median for the tone language group ($M = 0.25, 0.40$ for intonation language speakers; median = 0.45 for tone language group). No participants were identified as outliers.

A point of concern for interval discrimination trials in Study 2, which did not apply to Study 1, was that some participants generated negative H − FA difference scores. These scores may have resulted from confusion about the task, in that *no change* trials were most often labeled as including a change in interval size, and vice versa, for these participants. In order to address this possibility, we reanalyzed the data after transforming all negative difference scores to positive difference scores. The difference between groups was still significant after this transformation [$t(20) = 2.26, p < .05$]. Thus, the difference between groups reported above is not fully attributable to negative difference scores.

Perception and production combined. As with our analysis of Study 1, we addressed the perception–production relationship by correlating scores on the perception and production tasks. Absolute interval errors (averaged across trials for each individual) were significantly correlated with H − FA difference scores for interval discrimination when both samples were pooled together [$r(20) = -.68, p < .05$]. Within groups, this relationship was significant only among tone language speakers [$r(9) = -.59$] and intonation language speakers [$r(9) = -.70$]. Again, correlations between absolute note errors and H − FA difference scores for note discrimination were nonsignificant, both across and within language groups.

Speed and accuracy. Finally, we turn to the possibility of a speed–accuracy trade-off, which was a possible outcome for Study 2 but not Study 1. Participants in Study 2, unlike those in Study 1, were allowed to begin producing sequences whenever they were ready, whereas metronome clicks guided response times in Study 1. In addition, although instructions in Study 2 suggested adherence to the tempo of the target stimulus, participants may have differed in their rate of production. Thus, there are two ways

in which response timing in Study 2 may have led to a speed–accuracy trade-off: reaction time (response latency) and tempo (production rate). This issue is of particular concern because of recent data suggesting that some inaccurate singing is a by-product of tempo (Dalla Bella et al., 2007). Moreover, response latencies in singing have been linked to sequence complexity (Zurbriggen, Fontenot, & Meyer, 2006) and may similarly be linked to individual differences in skill.

We ran two ANOVAs, one on each measure of timing described above (reaction time, tempo), using the same design as we used to analyze production errors. Tone language speakers responded more slowly than intonation language speakers did [$F(1,20) = 4.30, p = .05$] and produced sequences more slowly [$F(1,20) = 7.81, p < .05$]. No interactions were found. However, an analysis of covariance conducted on absolute interval errors (which yielded a significant effect for the entire sample) revealed that the tone language advantage remained when reaction time and tempo are included as covariates [$F(1,19) = 7.56, p < .05$]. Correlations between timing and accuracy within each group (across participants) were nonsignificant for all pairings of pitch error measures (note/interval) with timing measures.

Discussion

Study 2 replicated the major findings of Study 1 with a more tightly controlled sample of (Mandarin) tone language speakers. Tone language speakers, in comparison with individuals who first learned an intonation language (in most cases, English), were more accurate at imitating musical pitch and were better at discriminating intervals. In addition, Study 2 (unlike Study 1) provided evidence for an advantage in the imitation of single pitches during production, although this advantage in the processing of absolute pitch was still not found in perception (note discrimination). Study 2 also included two bilingual speakers in the intonation language group. Their results suggest that the tone language advantage may not be due to bilingualism (all our tone language speakers were bilingual). Finally, the results of an expanded array of control tasks in Study 2 suggest that the tone language advantage is not linked to differences in vocal range or general cognitive ability. Study 2 also revealed that intonation language speakers imitated more quickly and responded sooner after the response cue than did the tone language speakers. This finding is reminiscent of a recent result from Dalla Bella et al. (2007), who found that differences between occasional (untrained) and professional singers were attributable to differences in production rate, which were faster for occasional singers. However, we found that the effect of language for interval errors (our most robust measure with respect to language groups) remained when timing effects were factored out. Given this result, and the fact that similar results were found regardless of whether timing was controlled for (cf. Study 1), we suggest that the results of Study 2 are unlikely to stem simply from a speed–accuracy trade-off.

GENERAL DISCUSSION

Both anecdotal reports and experimental studies have shown that people vary in their ability to process pitch, both in production and perception. We have reported the results of two studies that were designed to examine one possible source of individual differences in pitch-processing ability—namely, native language. Specifically, we hypothesized that acquiring a tone language during development would enhance an individual's ability to produce and perceive musical pitch due to the fine-grained pitch processing required by tone languages. The results bore out this prediction. Speakers of Asian tone languages (Mandarin, Vietnamese, and Cantonese) in our studies demonstrated greater accuracy in the imitation of pitch and in the discrimination of intervals than did native speakers of intonation languages (primarily English). The advantage for tone language speakers in production was contingent on complexity, suggesting that tone language speakers are better able than intonation language speakers to represent complex sequential relationships among successive pitches. By contrast, no advantage was seen in the discrimination of individual pitches, in vocal pitch range, or in a test of cognitive ability (the Wonderlic Personnel Inventory). These data suggest that the use of pitch to convey lexical information in one's native language facilitates the use of pitch in nonlinguistic contexts. The present data are therefore consistent with theoretical approaches that argue for neural integration of music and language (e.g., Brown, 2000), as opposed to models that assume specific modules for speech (or music) processing (e.g., Liberman & Mattingly, 1985).

We interpret the present results in light of the associations formed between pitch and lexical categories during language acquisition. In tone languages, pitch changes are correlated in a meaningful way with lexical items, leading to enhanced categorical perception for linguistic tones among speakers of tone languages (Stagray & Downs, 1993; Xu et al., 2006). In intonation languages, the link between pitch and meaning is more flexible and occurs more at the phrase level than at the lexical level. The fine-tuning of lexical pitch categories among tone language speakers then carries over into musical contexts, generally speaking. A possible mechanism for this transfer is *perceptual attunement* (cf. Gibson, 1963). During language acquisition, speakers become better able to perceive those characteristics of the stimulus array that are most informative to the task at hand. Although pitch is important to all language learners, it has a more specific role for tone language speakers, and thus those speakers may come to be more sensitive to the pitch dimension in general.

The present study focused on group differences in pitch processing as a function of linguistic background. Our previous study (Pfordresher & Brown, 2007) looked at individual differences in pitch accuracy among intonation language speakers alone and led us to characterize some individuals as "poor-pitch singers" (Welch, 1979) on the basis of their note errors during vocal imitation tasks. Using a signed note error of greater than ±100 cents as a standard, we found that 15% of the sample could be characterized as poor-pitch singers. It is instructive to examine the incidence of poor-pitch singers in the two linguistic groups analyzed in the present study. According to the ±100-cent standard, 8 singers from the two studies reported here would be classified as poor-pitch singers. This 17% figure is quite close to that reported by Pfordresher and Brown and by Dalla Bella et al. (2007). Although poor-pitch singers were clearly found in both linguistic groups, there was a trend toward greater numbers of them in the intonation language group than in the tone language group (5 vs. 3).

Another important comparison between our previous study and the present one relates to the relationship between perception and production. In our previous study, there was no relationship between perception and production, and poor-pitch singers were no more likely than accurate singers to be poor discriminators (see also Bradshaw & McHenry, 2005; Dalla Bella et al., 2007; cf. Loui, Guenther, Mathys, & Schlaug, 2008). We found the same result in the present studies: When examining the performance of poor-pitch singers in either linguistic group, we found no difference between their discrimination abilities and those of other participants. This was true for both note discrimination (H − FA, $M_{good\ singers}$ = .78, $M_{poor\ singers}$ = .75, p > .30) and interval discrimination ($M_{good\ singers}$ = .35, $M_{poor\ singers}$ = .21, p > .10). In contrast to this, the major point of difference between our previous study and this one was the *association* between perception and production seen in the tone language advantage, especially with regard to interval processing. How can we reconcile the present association with the previously established dissociation?

We argued in our previous article (Pfordresher & Brown, 2007) that the deficit underlying poor-pitch singing is based neither on purely sensory nor on purely motor factors but instead on a sensorimotor disruption. By this account, pitch accuracy during singing development requires the active engagement of sensorimotor mechanisms for pitch. If these mechanisms are underused during development, perception may remain intact, but pitch-matching mechanisms may become dissociated from it through a disruption of sensorimotor linkages. The tone language advantage—with its association between perception and production—may represent the opposite end of the spectrum: a use-dependent enhancement of audio–vocal linkages during the development of pitch processing. If the system for pitch processing is indeed shared between song and speech, lexical tone might be a parallel route to actual singing in developing pitch accuracy in music. Given the observed presence of inaccurate singers in our tone language sample, a reasonable but unexplored question becomes what the repercussions of deficient pitch-matching skills are (if any) for tone production. Future studies should examine whether a deficit in pitch imitation has consequences for tone language speakers in the context of linguistic production (e.g., intelligibility).

Moreover, the fact that certain tone language speakers did not perform well on our production tasks, or even the

perceptual discrimination tasks, has important implications for the present results. An inability to produce or imitate pitch contours in a tone language would certainly be detrimental for communication. However, there was no evidence that the tone language speakers who performed poorly on our musical tasks had trouble with speech communication; the method of Study 2 in particular controlled for that possibility. Thus, the implication is that, although the representation of linguistic pitch transfers to the representation of musical pitch, the two domains may not contain a fully shared representation. It may be that there is partial sharing or that two unshared representations interact. Another possibility, suggested recently (Patel, 2008), is that music and language share resources but not "representations." Applying this idea to the present data, tone languages may facilitate the use of resources for pitch processing in both music and language, but these resources may be used to support distinct representations of pitch. As such, certain tone language speakers may fail to do well on our musical tasks, because of problems with representation, whereas tone language speakers in general may show an advantage on musical tasks because of a more efficient use of resources.

As was mentioned in the introduction, recent theories have supported a proposition that resources shared by music and language may be specific to relative-pitch information (cf. Patel et al., 2008; Peretz & Coltheart, 2003). Some results in the present study do suggest such a link; nevertheless, we think it is premature to conclude that the tone language advantage is specific to relative pitch. Although our perceptual tasks showed no effect of linguistic background on note discrimination, recent research from another lab, but using the same type of stimuli, did find a tone language advantage for note discrimination (Stanley, Narayana, Pfordresher, & Wicha, 2008); that study involved electroencephalograph (EEG) recordings and therefore incorporated many more trials than ours did. Moreover, other research discussed earlier has documented the greater incidence of absolute-pitch labeling ability in tone language speakers who are musically trained (Deutsch et al., 2006). Thus, extended training (even within an experimental session) may alter the null effect of language background on note discrimination found here. With respect to production tasks, results from note errors differed across studies but pointed to the presence of a facilitation effect for the imitation of absolute pitch, although this was weaker than the effect for relative pitch. The fairest assessment of this and related results is to say that tone language speakers are generally better in the processing of musical pitch in a way that potentially generalizes to both relative and absolute pitch.

One difficulty in the design of the present study, which focused on an organismic variable (i.e., linguistic background), is the inability to eliminate factors other than the one presumed to cause the observed effects (cf. Schellenberg & Peretz, 2007). We cannot eliminate all possible alternative explanations, although our results argue against two: vocal range (as measured by the warm-up tasks) and general cognitive ability (as measured by the Wonderlic

Personnel Inventory). Related to this, at the present time we cannot determine whether group differences are due to linguistic environment or genetics. Our preferred explanation is framed in terms of environment. Nevertheless, recent findings suggest that genetic differences may exist between tone language speakers and speakers of intonation languages (Dediu & Ladd, 2007),[5] and it is possible that the facilitation observed in the present study is linked to such genetic differences.

Finally, we reflect on the potential significance of our imitation task. A tone language advantage was found for the imitation of novel sequences but not for the reproduction of "Happy Birthday" from long-term memory. We see this result as reflective of the way in which pitch processing is fine-tuned during the acquisition of a tone language. Infants are confronted with a vast array of novel sequences, and language acquisition is driven in part by the imitation of these sequences. As such, individual differences in the ability to reproduce pitch patterns, as a function of linguistic background, may be larger for the imitation of novel sequences than for well-learned sequences that can benefit from repeated performance. An important factor in the imitation of novel sequences is working memory, whereas working memory burdens are perhaps reduced for the reproduction of well-known melodies.[6] Thus, the more efficient use of pitch-processing resources, as argued earlier, may reduce working memory demands for tone language speakers. Indeed, this possibility explains one of the more prominent findings of the present article: the effect of complexity. Across tasks, the advantage for tone language speakers was present for more complex rather than for simpler tasks. In imitation, the advantage was found for the most complex sequence types (melody sequences). In perception, the advantage was seen in the more complex (difficult) interval discrimination task rather than in the simpler note discrimination task.

We have presented the first evidence (to our knowledge) of a joint benefit to both the production and perception of musical pitch among speakers of Asian tone languages relative to speakers of (European) intonation languages (primarily English). In doing so, we have identified one of a host of moderating variables that contribute to individual differences in the processing of musical pitch. The fact that the variable we identified is linguistic background has important theoretical implications insofar as it argues for pitch-processing resources that are shared across domains.

respondence concerning this article should be addressed to P. Q. Pfordresher, Department of Psychology, 355 Park Hall, University at Buffalo, Buffalo, NY 14260 (e-mail: pqp@buffalo.edu).

REFERENCES

Amir, O., Amir, N., & Kishon-Rabin, L. (2003). The effect of superior auditory skills on vocal accuracy. *Journal of the Acoustical Society of America*, **113**, 1102-1108. doi:10.1121/1.1536632

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain*, **125**, 238-251.

Boersma, P., & Weenink, D. (2008). Praat: Doing phonetics by computer (Version 5.0.25) [Computer program]. Retrieved May 31, 2008 from http://222.praat.org/.

Bradshaw, E., & McHenry, M. A. (2005). Pitch discrimination and pitch matching abilities of adults who sing inaccurately. *Journal of Voice*, **19**, 431-439.

Brown, S. (2000). The "musilanguage" model of music evolution. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 271-301). Cambridge, MA: MIT Press.

Burns, E. M., & Ward, W. D. (1978). Categorical perception— phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *Journal of the Acoustical Society of America*, **63**, 456-468.

Cruttenden, A. (1997). *Intonation* (2nd ed.). Cambridge: Cambridge University Press.

Cuddy, L. L., Balkwill, L.-L., Peretz, I., & Holden, R. R. (2005). Musical difficulties are rare: A study of "tone deafness" among university students. In G. Avanzini, L. Lopez, S. Koelsch, & M. Majno (Eds.), *The neurosciences and music II: From perception to performance* (Annals of the New York Academy of Sciences, Vol. 1060, pp. 311-324). New York: New York Academy of Sciences.

Dalla Bella, S., Giguère, J.-F., & Peretz, I. (2007). Singing proficiency in the general population. *Journal of the Acoustical Society of America*, **121**, 1182-1189.

Dediu, D., & Ladd, D. R. (2007). Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin. *Proceedings of the National Academy of Sciences*, **104**, 10944-10949.

Deliège, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl and Jackendoff's grouping preference rules. *Music Perception*, **4**, 325-360.

Deutsch, D., Henthorn, T., & Dolson, M. (2004). Absolute pitch, speech, and tone language: Some experiments and a proposed framework. *Music Perception*, **21**, 339-356.

Deutsch, D., Henthorn, T., Marvin, E., & Xu, H. (2006). Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period. *Journal of the Acoustical Society of America*, **119**, 719-722.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, **13**, 361-377.

Gandour, J., Wong, D., Hsieh, L., Weinzapfel, B., Van Lackner, D., & Hutchins, G. D. (2000). A crosslinguistic PET study of tone perception. *Journal of Cognitive Neuroscience*, **12**, 207-222.

Gandour, J., Wong, D., & Hutchins, G. (1998). Pitch processing in the human brain is influenced by language experience. *NeuroReport*, **9**, 2115-2119.

Gibson, E. J. (1963). Perceptual learning. *Annual Review of Psychology*, **14**, 29-56.

Hawkins, K. A., Faraone, S. V., Pepple, J. R., Seidman, L., & Tsuang, M. T. (1990). WAIS–R validation of the Wonderlic Personnel Test as a brief intelligence measure in a psychiatric sample. *Psychological Assessment*, **2**, 198-201.

Hickok, G., Buchsbaum, G., Humphries, C., & Muftuler, T. (2003). Auditory–motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience*, **15**, 673-682.

Holleran, S., Jones, M. R., & Butler, D. (1995). Perceiving implied harmony: The influence of melodic and harmonic context. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 737-753.

Jones, M. R. (1987). Dynamic pattern structure in music: Recent theory and research. *Perception & Psychophysics*, **41**, 621-634.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.

Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception & Psychophysics*, **56**, 414-423.

Levitin, D. J., & Rogers, S. E. (2005). Absolute pitch: Perception, coding, and controversies. *Trends in Cognitive Sciences*, **9**, 26-33.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, **74**, 431-461.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revisited. *Cognition*, **21**, 1-36.

Loui, P., Guenther, F., Mathys, C., & Schlaug, G. (2008). Action–perception mismatch in tone-deafness. *Current Biology*, **18**, R331-R332.

McCann, J., & Peppé, S. (2003). Prosody in autism spectrum disorders: A critical review. *International Journal of Language & Communication Disorders*, **38**, 325-350.

Milenkovic, P. H. (2001). *TF32* [Computer software and manual]. Retrieved January 7, 2005, from http://userpages.chorus.net/cspeech.

Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The implication–realization model*. Chicago: University of Chicago Press.

Patel, A. D. (2003). Language, music, syntax, and the brain. *Nature Neuroscience*, **6**, 674-681.

Patel, A. D. (2008). *Music, language, and the brain*. New York: Oxford University Press.

Patel, A. D., Wong, M., Foxton, J., Lochy, A., & Peretz, I. (2008). Speech intonation perception deficits in musical tone deafness (congenital amusia). *Music Perception*, **25**, 357-368.

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

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, **6**, 688-691.

Peretz, I., & Zatorre, R. J. (2005). Brain organization for music processing. *Annual Review of Psychology*, **56**, 89-114.

Pfordresher, P. Q., & Brown, S. (2007). Poor-pitch singing in the absence of "tone deafness." *Music Perception*, **25**, 95-115.

Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., & Lang, J. M. (1994). On the perceptual organization of speech. *Psychological Review*, **101**, 129-156.

Schellenberg, E. G. (2003). Does exposure to music have beneficial side effects? In I. Peretz & R. J. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 430-448). New York: Oxford University Press.

Schellenberg, E. G., & Peretz, I. (2007). Music, language and cognition: Unresolved issues. *Trends in Cognitive Sciences*, **12**, 45-46.

Schellenberg, E. G., & Trehub, S. E. (2003). Good pitch memory is widespread. *Psychological Science*, **14**, 262-266.

Schmidt, R. A., & Lee, T. D. (1999). *Motor control and learning: A behavioral emphasis*. Champaign, IL: Human Kinetics.

Sloboda, J. A., Wise, K. J., & Peretz, I. (2005). Quantifying tone deafness in the general population. In G. Avanzini, L. Lopez, S. Koelsch, & M. Majno (Eds.), *The neurosciences and music II: From perception to performance* (Annals of the New York Academy of Sciences, Vol. 1060, pp. 255-261). New York: New York Academy of Sciences.

Smith, J. D. (1997). The place of musical novices in music science. *Music Perception*, **14**, 227-262.

Smith, J. D., Kemler Nelson, D. G., Grohskopf, L. A., & Appleton, T. (1994). What child is this? What interval was that? Familiar tunes and music perception in novice listeners. *Cognition*, **52**, 23-54.

Stagray, J. R., & Downs, D. (1993). Differential sensitivity for frequency among speakers of a tone and a nontone language. *Journal of Chinese Linguistics*, **21**, 143-163.

Stanley, E., Narayana, S., Pfordresher, P. Q., & Wicha, N. (2008). Advantage of tonal language speaking on pitch perception. *Journal of Cognitive Neuroscience*, **20** (Suppl. 1, p. 232).

Takeuchi, A. H., & Hulse, S. H. (1993). Absolute pitch. *Psychological Bulletin*, **113**, 345-361.

Todd, R., Boltz, M. G., & Jones, M. R. (1989). The MIDILAB auditory research system. *Psychomusicology*, **8**, 17-30.

TROUT, J. D. (2001). The biological basis of speech: What to infer from talking to the animals. *Psychological Review*, **108**, 523-549.

WARD, W. D. (1999). Absolute pitch. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 265-298). San Diego: Academic Press.

WELCH, G. F. (1979). Poor pitch singing: A review of the literature. *Psychology of Music*, **7**, 50-58.

WENNERSTROM, A. (2001). *The music of everyday speech: Prosody and discourse analysis*. New York: Oxford University Press.

WISE, K. J., & SLOBODA, J. A. (2008). Establishing an empirical profile of self-defined "tone deafness": Perception, singing performance and self-assessment. *Musicae Scientiae*, **12**, 3-23.

WONG, P. C. M., PARSONS, L. M., MARTINEZ, M., & DIEHL, R. L. (2004). The role of the insular cortex in pitch pattern perception: The effect of linguistic contexts. *Journal of Neuroscience*, **24**, 9153-9160.

XU, Y., GANDOUR, J. T., & FRANCIS, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *Journal of the Acoustical Society of America*, **120**, 1063-1074.

YIP, M. (2002). *Tone*. Cambridge: Cambridge University Press.

ZURBRIGGEN, E. L., FONTENOT, D. L., & MEYER, D. E. (2006). Representation and execution of vocal motor programs for expert singing of tonal melodies. *Journal of Experimental Psychology: Human Perception & Performance*, **32**, 944-963.

NOTES

1. The accuracy of participants who received payment did not differ from those who participated for course credit.

2. One tone language speaker did not know the song.

3. Study 1 did not include an independent measure of vocal range, a shortcoming that was addressed in Study 2.

4. As in Study 1, those who were paid did not perform better than those who volunteered for course credit. In fact, those who were paid performed nonsignificantly worse in all tasks.

5. We thank Isabelle Peretz for informing us about this article.

6. We thank an anonymous reviewer for pointing this out.