The impact of musical training and tone language experience on talker identification

Xin Xie
Department of Psychology, University of Connecticut, 406 Babbidge Road, Unit 1020, Storrs, Connecticut 06269

Emily Myers[a)]
Department of Speech, Language, and Hearing Sciences, University of Connecticut, 850 Bolton Road, Unit 1085, Storrs, Connecticut 06269

(Received 30 June 2014; revised 19 November 2014; accepted 30 November 2014)

Listeners can use pitch changes in speech to identify talkers. Individuals exhibit large variability in sensitivity to pitch and in accuracy perceiving talker identity. In particular, people who have musical training or long-term tone language use are found to have enhanced pitch perception. In the present study, the influence of pitch experience on talker identification was investigated as listeners identified talkers in native language as well as non-native languages. Experiment 1 was designed to explore the influence of pitch experience on talker identification in two groups of individuals with potential advantages for pitch processing: musicians and tone language speakers. Experiment 2 further investigated individual differences in pitch processing and the contribution to talker identification by testing a mediation model. Cumulatively, the results suggested that (a) musical training confers an advantage for talker identification, supporting a shared resources hypothesis regarding music and language and (b) linguistic use of lexical tones also increases accuracy in hearing talker identity. Importantly, these two types of hearing experience enhance talker identification by sharpening pitch perception skills in a domain-general manner.
[http://dx.doi.org/10.1121/1.4904699]

[TCB] Pages: 419–432

I. INTRODUCTION

Listeners skillfully extract information in the speech signal to infer talker identity: For instance, listeners can quickly recognize the voice of a familiar talker on the phone. However, the mechanisms that underlie such remarkable talker identification ability remain relatively unclear (Belin *et al.*, 2011). Early studies on talker identification focused on uncovering acoustic correlates of talker identity such as pitch, hoarseness, voice quality, etc., (e.g., Gelfer, 1988). Generally, studies of talker identification employed two complementary approaches: Linking natural, unmodified acoustic properties of talkers to their perceived similarity/distinctiveness or examining the effect of acoustic alterations on talker identification performance. Studies using the first approach exposed listeners to natural productions from a large number of talkers and asked them to distinguish or to rate similarities among talkers. Acoustic dimensions that were different in talkers who were rated as distinct but were shared by talkers who were rated as highly similar are considered important for talker identification. Studies using the second approach manipulated certain acoustic cues while controlling others to evaluate the weighting of changed parameters in talker identification (e.g., Remez *et al.*, 1997). Voice pitch, the auditory perception of the rate of vocal fold vibration (the fundamental frequency or F_0), has emerged as an important acoustic cue of talker identity that was consistently used by listeners across such studies (see Creel and Bregman, 2011 for a review). Moreover, multidimensional scaling studies also show that F_0 variation is the primary parameter used in differentiating between speakers (e.g., Baumann and Belin, 2010). It is important to note that although *pitch* is a reliable cue for vocal identity, other cues (e.g., vocal timbre, speaking rate) that reveal indexical characteristics can be used for talker identification. In addition, there is strong evidence that language-dependent use of certain acoustic-phonetic properties (e.g., formants, voice onset time or VOT) also helps listeners to identify talkers. For instance, formants in vowels inform listeners of talker gender (Remez *et al.*, 1997). Listeners can also use VOT of consonants to cue talker identity after enough training (Francis and Driscoll, 2006). Importantly, these acoustic-phonetic properties may be tightly linked to knowledge of the language being spoken, whereas pitch may also be used when this information is unavailable, such as when listeners are hearing voices speaking an unfamiliar language.

People make use of pitch variation not only to recognize human voices but also in a variety of everyday experiences: From humming a song to interpreting the emotion in a spoken message (Juslin and Laukka, 2003). Nevertheless, individual listeners markedly differ in the ability to perceive variations in pitch and to encode this variability in memory (e.g., Pfordresher and Brown, 2009; Gaab and Schlaug, 2003). Recent studies focusing on group characteristics have found that both musical and linguistic experience with pitch use can enhance pitch perception. Perhaps unsurprisingly, musicians show superior performance compared to non-musicians in perceptual tasks such as pitch discrimination and pitch change detection (Tervaniemi *et al.*, 2005;

a)Author to whom correspondence should be addressed. Electronic mail: emily.myers@uconn.edu

Bidelman *et al.*, 2013). Importantly, training with musical pitch appears to transfer to linguistic contexts: Musicians also show better performance (higher accuracy and faster responses) when discriminating lexical tones of an unfamiliar tone language (Burnham *et al.*, 2014) and greater sensitivity to prosodic pitch changes in spoken sentences (e.g., Deguchi *et al.*, 2012). This transfer appears to be bidirectional: Linguistic use of tones in language likewise equips tone language users with enhanced pitch sensitivity for both linguistic pitch in lexical tones (Krishnan *et al.*, 2009) and musical pitch (Pfordresher and Brown, 2009).

Notably, individuals also vary widely in their ability to identify talkers. Schmidt-Nielsen and Crystal (1998) examined listener performance in a same/different talker discrimination task. In this study, accuracy ranged from 52% to 85% across all 65 normal-hearing individual listeners. Substantial individual variability is also observed in other studies investigating voice recognition (e.g., Winters *et al.*, 2008). Such individual differences pose a special problem for voice lineups in forensic cases: Evidently, earwitnesses vary tremendously in their competency to identify speakers in voice parades. Trained personnel such as phoneticians, speech pathologists, or social workers are more trusted in the courtroom than others. Experts on forensic speaker identification have suggested that layperson earwitnesses should not be used at all due to the uncertainty in their speaker identification accuracy (e.g., Künzel, 1994). Nevertheless, apart from professional training on talker identification, little is known about what leads to high or low performance in individual listeners. Cochlear implant users have documented difficulties in speaker recognition (e.g., Cleary and Pisoni, 2002). Given that this population may, among other things, experience poorer pitch resolution, it is possible that pitch abilities contribute to difficulties in speaker recognition in CI users. Yet few studies have examined the core perceptual or cognitive factors that contribute to individual differences in talker identification in the typical population (cf. Bregman and Creel, 2014). Given that pitch constitutes a significant acoustic component of voice identity, we theorize that there is a close link between individual abilities in pitch processing and varied talker identification skills.

It is widely acknowledged that pitch perception involves auditory processing at two structural levels: global and local (e.g., Peretz, 1990; Sanders and Poeppel, 2007). At the global level, listeners need to process the contour patterns of pitch changes; at the local level, listeners perceive the absolute pitches or precise intervals that make up a contour. In music, musical intervals and melodic contours are different elements. In language, pitch conveys rich information about the structure of speech: word stress, sentence prosody, speaker emotion, to name a few (e.g., Juslin and Laukka, 2003). Additionally in tone languages, pitch also carries lexical information. For example, in the tone language Mandarin, the same syllable /ma/ can mean *mother*, *linen*, *horse,* or *scold* when implemented with different pitch contours. In a broad sense, pitch perception in speech involves listeners' sensitivity to pitch changes over different temporal scales. Some linguistic distinctions require attention to local changes in pitch height, as do lexical tones in many African tone languages (e.g., Yoruba); some are expressed via short contours, as lexical tones in some Asian tone languages (e.g., Mandarin). Over even longer intervals, global pitch perception is important for speech prosody such as word stress or sentence intonation.

However, despite the clear distinction between global and local processing in pitch perception studies, pitch is usually referred to as a single acoustic/perceptual element in studies on human voice recognition. It is unclear *how* exactly pitch is used in talker identification. On the one hand, idiosyncratic prosodic changes, and specifically dynamics of the F_0 contour, can be used for discriminating speakers (Mary and Yegnanarayana, 2008). On the other hand, the absolute pitch height also reveals talker information in the sense that each talker has a stereotypic F_0 that is partially a product of his/her laryngeal anatomy. For example, by manipulating the pitch height of synthetic speech, researchers can change listeners' perception of the number of talkers in a dialogue (Magnuson and Nusbaum, 2007). An individual difference approach provides a potent tool to understand the role of pitch perception in talker identification; and specifically, it would help tease apart various uses of a pitch processing system in talker identification. Previous studies have shown that global and local processing of pitch are dissociable when related to other linguistic skills, such as reading ability (e.g., Foxton *et al.*, 2003). By linking individual differences in global versus local pitch perception with listener variability in identifying talkers, we will understand better how pitch processing contributes to talker identification.

In the current study, we built on the overarching hypothesis that pitch processing abilities are related to talker identification by investigating both group-level performance (experiment 1) and individual differences (experiment 2) in talker identification. In experiment 1, we investigated whether tone language users and amateur musicians, two groups hypothesized to have advantages in pitch processing, have better talker identification than speakers of languages without lexical tones and non-musicians. In addition, we examined the proposed perceptual benefits in different linguistic contexts: in one's native versus unfamiliar languages. We hypothesize that (1) tone language users (Mandarin listeners) will have enhanced talker identification compared to speakers of non-tone languages (English listeners), controlling for language familiarity and experience with musical training. (2) Musicians will have better talker identification in general than non-musicians. Experiment 2 further allowed us to explore individual pitch perception abilities and their relation to talker identification. Specifically, we used a mediation analysis approach (Baron and Kenny, 1986) to test the hypothesis that the observed superiority in speaker identification performance is mediated by pitch processing ability of individual listeners. We used both local pitch and global pitch discrimination tasks to examine whether either or both types of pitch perception are good predictors of talker identification performance.

II. EXPERIMENT 1

In experiment 1, we tested the hypothesis that individuals with musical training or tone language experience will

have enhanced talker identification performance compared to those who do not have such experience. It is important to note that the accuracy of talker identification is sensitive to the shared language background between the speaker and the listener. Namely, listeners have more difficulty identifying talkers in unfamiliar languages compared to their native language, a phenomenon known as the *language familiarity effect* (Perrachione *et al.*, 2011; Perrachione *et al.*, 2009). This advantage is assumed to arise because native speakers can access talker-idiosyncratic phonetic variation in their native language (e.g., Remez *et al.*, 1997; Winters *et al.*, 2008). For this reason, we compared native-Mandarin versus native-English listeners' performance in identifying Mandarin, English, and Spanish talkers. First, having multiple language conditions will help to illuminate whether musically trained individuals and tone language speakers have better judgment in talker identity in general, regardless of the degree of access to the linguistic content of the utterance. In addition, the comparison across language conditions will help to show whether the group advantages, if any, are affected by language-dependent factors. Given that listeners are hypothesized to use language-specific cues more strongly in their native language, in the absence of those cues (the non-native listening conditions), we predict that perceptual benefits originating from enhanced pitch perception will be more pronounced. Spanish is a foreign language to both listener groups, and this language condition should provide a situation in which we can directly test whether tone language users (Mandarin listeners) outperform non-tone language users (English listeners) in talker identification, without the confounding effect of language familiarity. Within the native-English group, we compared talker identification accuracy in listeners with and without musical training.

A. Methods

1. Participants

Two groups of listeners were recruited from the University of Connecticut to participate in the study. A self-report questionnaire was used to collect information about listeners' language and musical background: age of acquisition (AoA) for L2 (if applicable), the starting and finishing ages of any musical training, including the name of musical instruments/vocal training. We also asked participants to report the type of settings (e.g., recitals, private lessons, etc.) in which they tended to practice the instruments and the frequency of such practice (hours/day, and times/week). Previous studies differed in the cut-off criterion used to define musicianship in young adults. The inclusion criteria for musicians range from 6 to 10 years of musical training (e.g., Bidelman *et al.*, 2011; Strait *et al.*, 2010; Wayland *et al.*, 2010). We took the lower end and defined "musicians" as amateur instrumentalists or vocalists with at least 6 years of continuous formally instructed musical training throughout their lifetime (Wayland *et al.*, 2010; Chan *et al.*, 1998). Note that this relatively liberal criterion is more likely to produce a result that goes against our prediction by finding no differences between musicians and non-musicians. Non-musicians were defined as individuals who had received less than 1 year of formal vocal training or training on any musical instrument(s). Given that pitch perception is the focus of the current investigation, experience with percussion instruments was not included. After excluding participants who scored at or below chance in the talker identification task in their native language ($n = 10$, five Mandarin and five English), 36 native-English listeners, and 25 native-Mandarin listeners who speak English as L2 were included for data analyses. The English group was divided into 10 musicians (years of formal training: $M = 10.50$, SD $= 2.80$; age of onset of musical training: $M = 7.60$, SD $= 3.10$) and 26 non-musicians; similar to the English non-musicians, all Mandarin speakers had less than 1 year of musical training with no significant difference between the English Non-Musician and Mandarin Non-Musician groups on years of musical training [$t(49) = 1.846$, $p = 0.07$]. All English listeners were naive to Mandarin; however, 24 of them had studied Spanish at school at some point (age of acquisition: $M = 11.25$, SD $= 3.14$; length of study: $M = 5.46$, SD $= 2.96$) although none were fluent in the language. Critically, among the English Musician group, seven participants (of 10) had learned Spanish (age of acquisition: $M = 10.71$, SD $= 2.63$; length of study: $M = 5.00$, SD $= 2.65$); among the English Non-Musician group, 17 participants (of 26) had learned Spanish (age of acquisition: $M = 11.47$, SD $= 3.37$; length of study: $M = 5.65$, SD $= 3.14$). The English Musician group did not differ from the English Non-Musician group in terms of the onset of Spanish learning (calculated for those who had Spanish experience), $t(22) = 0.528$, $p = 0.60$; or the length of prior Spanish experience, $t(34) = 0.144$, $p = 0.89$. All Mandarin listeners were naive to Spanish but were able to use English to conduct daily conversation. Mandarin listeners were late bilinguals who learned English in classroom setting in Mainland China; their average age of acquisition was 10.16 (SD $= 2.88$) years old, and their age of arrival in the U.S. was 22.46 (SD $= 3.37$) years old. All participants were undergraduates or graduate students at the University of Connecticut and received academic credits or monetary rewards for participation. None reported a hearing or visual disorder. No listener reported prior familiarity with any of the voices in the listening experiment. A written informed consent was obtained from every participant in accordance with the guidelines of the University of Connecticut IRB.

2. Stimuli

Stimuli for the talker identification task consisted of recordings of 10 sentences in each language condition (Mandarin, Spanish, and American English) (see the Appendix). Five male native speakers of each language read all 10 sentences in that language condition. The five speakers in each language condition and the 10 sentences were selected from a larger sample of speakers and sentences that were originally recorded to control for sentence duration, variation in fundamental frequency and noticeable idiosyncratic accent. Sentences were created by the experimenter and checked for naturalness by two native speakers of each language. All 15 speakers were perceived as having no discernible idiosyncratic

talker characteristics (e.g., unusual prosodic patterns or vocal quality) by colleagues of the experimenters. Speakers were instructed to read naturally as if talking to a friend. No speaker participated in more than one language condition or in the listening experiment. Sentence duration and within-condition variation of F_0 among the five talkers were controlled across conditions such that there were no significant differences between language conditions on these measures. Recordings were made in a sound-proof room and then digitally sampled at 22.05 kHz and normalized for RMS amplitude to 70 dB sound pressure level (SPL). Five sentences in each language were arbitrarily designated as training sentences and the remaining five as testing sentences.

3. Procedure

Each participant performed the talker identification task in all three language conditions (English, Mandarin, and Spanish). The task was blocked by language condition with the order of language counterbalanced across participants and within each listener group. Each block consisted of a familiarization phase, a practice phase and a generalization phase in the same language.

a. Familiarization phase. Participants were seated in front of a computer monitor with auditory stimuli presented over headphones. On each trial, each participant heard one of the five training sentences read by one of the five speakers while a number designating that speaker's voice (1, 2, 3, 4, or 5) appeared on the computer monitor. Then the participant typed in the number that they saw. This simple procedure was intended to keep participants' attention focused on the task and to familiarize them with the voices. The presentation order was blocked by sentences. That is, during 10 consecutive trials, participants heard the same sentence read by all five speakers with two repetitions from each speaker in a row. This procedure was repeated until listeners heard all five voices reading all five training sentences. The total number of sentences they heard was 2 times × 5 sentences × 5 talkers = 50 trials.

b. Practice phase. After familiarization, participants learned to identify the voice of each talker, with feedback. After one of the five speakers read the training sentence, the participant tried to identify the voice by entering 1, 2, 3, 4, or 5 on the keyboard. Feedback was given to the participant, and the correct talker was indicated if the response was incorrect. Then the participant advanced to the next trial by pressing a button on the keyboard. During the identification session for each sentence, three tokens of each of all five speakers reading five trained sentences were presented in randomized order, resulting in a total of 75 trials.

c. Generalization phase. After practicing, participants were tested on their ability to identify the voices from five novel sentences without any feedback. The procedure was the same as in the practice phase. This phase helped to assess the generalizability of their talker identification.

After completing one language condition, participants took a short break and repeated the same task in another language condition. Upon completion of all listening tasks,[1] participants were asked to fill out a survey on their language and musical background.

B. Results

Listeners' accuracy in identifying talkers in the generalization phase of the talker identification task was examined. We compared the groups' accuracy using a 3 between-subject (group: English Non-Musicians, English Musicians, and Mandarin Non-Musicians) × 3 within-subject (language condition: English, Mandarin or Spanish voices) analysis of variance (ANOVA). The dependent measure was the percentage of correct identifications of talkers during the generalization phase. The ANOVA revealed a significant main effect of group, $F(2,58) = 4.848$, $p = 0.011$, $\eta_p^2 = 0.143$. Planned comparisons across language conditions were conducted to test our main hypothesis; as predicted, English Non-Musicians had significantly lower accuracy in talker identification than English Musicians and Mandarin Non-Musicians together ($p = 0.003$); between the two groups (English Musicians and Mandarin Non-Musicians) who were hypothesized to have better talker identification, no significant difference was found ($p = 0.20$). The ANOVA also revealed a significant effect of language condition, $F(2,116) = 17.129$, $p < 0.001$, $\eta_p^2 = 0.228$. Across all three groups, listeners had poorer performance with the Spanish talkers ($M = 0.52$, SD = 0.14) and had relatively higher accuracy with English talkers ($M = 0.64$, SD = 0.18) and Mandarin talkers ($M = 0.60$, SD = 0.18). Note that this result was driven by the fact that Spanish is a non-native language for every participant. There was a significant interaction between group and language condition, $F(4,116) = 30.386$, $p < 0.001$, $\eta_p^2 = 0.512$. We further examined this interaction by conducting three omnibus ANOVAs on English, Mandarin, and Spanish listening conditions, separately (Fig. 1). When listening to English sentences, we found a main effect of group, $F(2,58) = 9.706$, $p < 0.001$, $\eta_p^2 = 0.251$. *Post hoc* comparisons[2] revealed superior performance among English Musicians and Non-Musicians relative to Mandarin listeners (English Musicians = English Non-Musicians > Mandarin Non-Musicians, $p < 0.001$). Similarly,

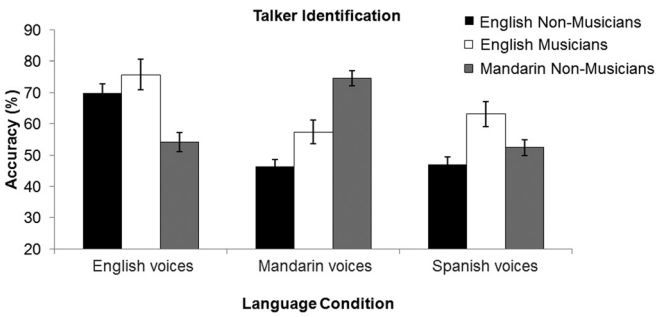

FIG. 1. Percentage of trials correctly identified by English Non-Musicians, English Musicians, and Mandarin Non-Musicians in the generalization phase in each language condition: English, Mandarin, and Spanish (experiment 1). Error bars indicate ±1 SEM. Significant difference is represented by ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.

for Mandarin talkers, a main effect of group was found, $F(2,58) = 36.085$, $p < 0.001$, $\eta_p^2 = 0.554$. *Post hoc* comparisons indicated that the main effect was driven by Mandarin listeners' higher accuracy than English listeners regardless of the English listeners' musical training (p's < 0.001). These results are a direct replication of the language familiarity effect shown in previous studies in that both language groups had enhanced performance in their own native language condition (e.g., Perrachione *et al.*, 2009). However, the important finding here is that English Musicians outperformed English Non-Musicians in an unfamiliar language, Mandarin [$t(34) = 2.737$, $p = 0.01$]. Again, when listening to Spanish talkers, we observed a significant group effect [$F(2,58) = 6.006$, $p < 0.005$, $\eta_p^2 = 0.172$]. *Post hoc* comparisons revealed that English Musicians again outperformed English Non-Musicians ($p < 0.005$) while Mandarin Non-Musicians did not differ from English Musicians ($p = 0.08$) or English Non-Musicians ($p = 0.37$).

C. Discussion

This experiment was designed to investigate whether musical and/or linguistic experience with pitch processing increases talker identification competency. Our hypothesis that musicians are better at talker identification than non-musicians was clearly confirmed in the English listener group. In particular, the benefit was found in the two unfamiliar language conditions (Mandarin and Spanish) but not in the native language condition. This effect is perhaps driven by the fact that listeners flexibly make use of a combination of different cues whenever they are accessible. In an unfamiliar language, listeners may rely more on low-level properties of the stimulus such as pitch variation; and as such, experience with musical pitch may only benefit listeners under such conditions. These results were consistent with previous findings (e.g., Winters *et al.*, 2008; Perrachione *et al.*, 2011) showing that listeners rely on language-dependent information when the speech is comprehensible but are forced to rely on language-independent indexical information in speakers' voices (e.g., their vocal pitch) when listening to speech in unfamiliar languages. It is important to note that the musician and non-musician groups were both naive to Mandarin and were equated on prior Spanish experience. Given similar language exposure between groups, we found it unlikely that the benefit of musicianship in the unfamiliar language conditions is due to musicians' linguistic competence with these languages.

Meanwhile, we also found that as tone language users, Mandarin non-musicians outperformed English non-musicians at this task overall when collapsing across all language conditions; and they did not differ from English musicians significantly. This finding is similar to a pattern reported (but not statistically tested) in a study by Perrachione and Wong (2007). Note that this finding alone cannot be taken as solid evidence showing tone language users' higher proficiency in talker identification. An alternative explanation could attribute the results to Mandarin listeners' knowledge of English, which might have potentially increased their perceptual accuracy in identifying English speakers and contributed to the overall better performance across language conditions. To clarify the situation, we conducted paired t-tests to compare Mandarin listeners' performance in each language condition. Given that Mandarin listeners in this study have some knowledge of English but no experience with Spanish, we would predict that they would show higher performance on English than Spanish if language familiarity is driving the effect. While Mandarin listeners showed the predicted higher accuracy in their native language (Mandarin: $M = 0.75$, SD $= 0.13$; Mandarin $>$ English $=$ Spanish, $p < 0.001$), there was no difference between the English condition ($M = 0.54$, SD $= 0.14$) and Spanish condition ($M = 0.52$, SD $= 0.11$). Thus the language knowledge of English among Mandarin listeners is unlikely the reason that they have outperformed the English non-musicians. Furthermore, when perceiving Spanish speakers, English musicians had significantly better performance than English non-musicians; meanwhile, Mandarin listeners' scores were intermediate between English musicians and non-musicians, although the differences between them did not achieve significance. Thus despite the finding that the perceptual benefit of long-term use of lexical tones was not as large as that elicited by extensive musical training, it is suggestive of a positive influence on talker identification.

Of note, all Mandarin listeners in this study were late bilinguals. It is therefore possible that the advantages we see in this group arise not from native language experience speaking a tone language but from potential perceptual advantages that might arise from being a bilingual (Krizman *et al.*, 2012). It is worth noting that a significant proportion of our English listeners also reported similar late foreign language experience with comparable starting age of acquisition and classroom learning environment, although we did not have sufficient information to assess the actual quality of learning in the two populations. The potential mediating effect of bilingualism will be addressed more fully in the general discussion.

Taken together, these results provided evidence to support our predictions. The perceptual benefit gained from musical training was clearly demonstrated in unfamiliar languages. However, unambiguous evidence of tone language use enhancing voice perception is relatively elusive—in particular, it remains possible that an advantage for non-native talker identification is evident in Mandarin listeners not due to their pitch abilities but rather because of other factors such as motivation, IQ, or degree of bilingualism. In experiment 2, we sought to replicate the findings of experiment 1 on musical training and more directly test the prediction regarding tone language experience, using a larger sample. In addition, we explicitly tested the prediction that individual differences in talker identification are mediated by listeners' pitch expertise, which can be promoted either by musical training or tone language use.

III. EXPERIMENT 2

The goal of experiment 2 was threefold. First, we wanted to further investigate the hypotheses that (a) musical and/or linguistic experience improves talker

identification and (b) the improvement originates from enhanced pitch-processing abilities. Although it is widely reported that musicians benefit from their pitch processing skills in multiple linguistic tasks (e.g., Wong and Perrachione, 2007; Bidelman *et al.,* 2013), enhanced pitch processing in musicians has never been directly associated with better talker identification (cf. Bregman and Creel, 2014). Similarly, despite the fact that cross-domain benefit of pitch processing for tone language users has been found in music and language studies, we do not know if speaking a tone language also makes one more sensitive to talker information in speech. In experiment 1, we had relatively unbalanced number of musicians versus non-musicians; experiment 2 addressed the questions with a more extensive sample. If musicians/tone language speakers show better performance in the pitch perception tasks and likewise show an advantage in talker identification tasks, we can substantiate the finding in experiment 1 that better musical and/or linguistic pitch processing may confer an advantage for talker identification. Second, on the individual level, if we can link individual performance in pitch perception with performance on talker identification tasks, we can corroborate the hypothesis that enhanced pitch processing is beneficial in the context of talker identification across individuals. Furthermore, we investigated two different types of pitch perception: local perception of pitch height changes and global perception of pitch contours, using a paradigm modeled after Foxton *et al.* (2003). Local versus global pitch perception tasks have been used to tap into auditory processing differences linked with hemispheric lateralization and are empirically dissociable in developmental disorders such as dyslexia (Foxton *et al.,* 2003). Both local and global pitch could be potentially useful in talker identification. F_0 height in low or high voices (local) could be a generalized cue to the standard pitch of a talker's voice. F_0 contour (global) may be used in cueing speech prosody, word stress, and other linguistic dimensions and thus can be exploited by listeners to capture talker-specific characteristics. In the current study, we examined local versus global pitch pattern perception in connection with their usage in talker identification in normal adults. We tested two language groups: Native-English (non-tone language) and native-Mandarin (tone language) speaking listeners. Within each group, we compared the performance between individuals varying in musical expertise.

A. Methods

1. Participants

A new sample of native American-English-speaking listeners ($n = 76$) and native Mandarin-speaking listeners ($n = 60$) with no known hearing or visual disorder was recruited from the University of Connecticut to participate in the study. Each participant completed the questionnaire on language and music background. We retained the same inclusion criteria for musicians (now labeled as "Extensive Training") and non-musicians as in experiment 1 but also included individuals who had intermediate musical training to fully capture the individual differences. Participants who had more than 1 year but fewer than 6 years of musical training at any point in their lives were categorized as musicians with minimal training (labeled as "Minimal Training").[3] One English participant and one Mandarin participant who failed to achieve above-chance performance in their native language condition for the talker identification task were excluded for further analysis. A total of 134 participants were then analyzed for talker identification and pitch perception: The English group consisted of 20 Extensive Training (years of musical training: $M = 9.20$, SD $= 2.86$), 23 Minimal Training ($M = 3.08$, SD $= 1.56$), and 32 Non-Musicians; the Mandarin group consisted of 14 Extensive Training ($M = 8.14$, SD $= 1.91$), 14 Minimal Training ($M = 2.68$, SD $= 1.61$), and 31 Non-Musicians. Overall, English listeners and Mandarin listeners were matched on musical training, $t(132) = 1.236$, $p = 0.22$. As in experiment 1, Mandarin listeners were late bilinguals who learned English at school in mainland China; their average age of acquisition was 10.20 (SD $= 2.86$) years old, and their age of arrival in the U.S. was 22.57 (SD $= 4.32$) years old. None of the English listeners had learned Mandarin before, nor did they have any regular exposure to it. Participants received academic credits or monetary rewards for participation. A written informed consent was obtained from every participant.

2. Stimuli

Stimuli for the talker identification task in the English and Mandarin conditions were the same as in experiment 1. Stimuli for both global and local pitch perception tasks were created following the paradigm established by Foxton *et al.* (2003). Each task contained 40 pairs of pure tone sequences (20 same, 20 different), presented at an ISI of 1s. Each sequence consisted of six pure tones (250 ms duration, 20 ms gating window), with pitches taken from an atonal octave, equally spaced by seven logarithmic steps. Starting pitches varied from 250 to 354 Hz. For the local pitch task, different trials were created by replacing a random note (avoiding the first and last notes), but keeping all other pitches the same between the two sequences. For the global pitch task, different trials were created by first shifting the whole pitch contour by a fixed interval and then replacing a random note, resulting in a violation of the original contour. That is, "same" trials contained sequences that were shifted in pitch but with contour maintained to prevent the use of absolute pitch cues in this task. The altered notes were always two notes higher or lower than the original tone (Fig. 2). Every sequence expanded over exactly five notes.

3. Procedure

All participants participated in the talker identification task first. The procedure was the same as in experiment 1. Following this task, participants completed the pitch perception tasks.[4]

In the pitch perception task, participants were instructed to make a same/different judgment for the two sequences in each trial and press "s" or "d" on the keyboard for response. In the local pitch task, participants were instructed to respond with "s" if there was no note change at all and two

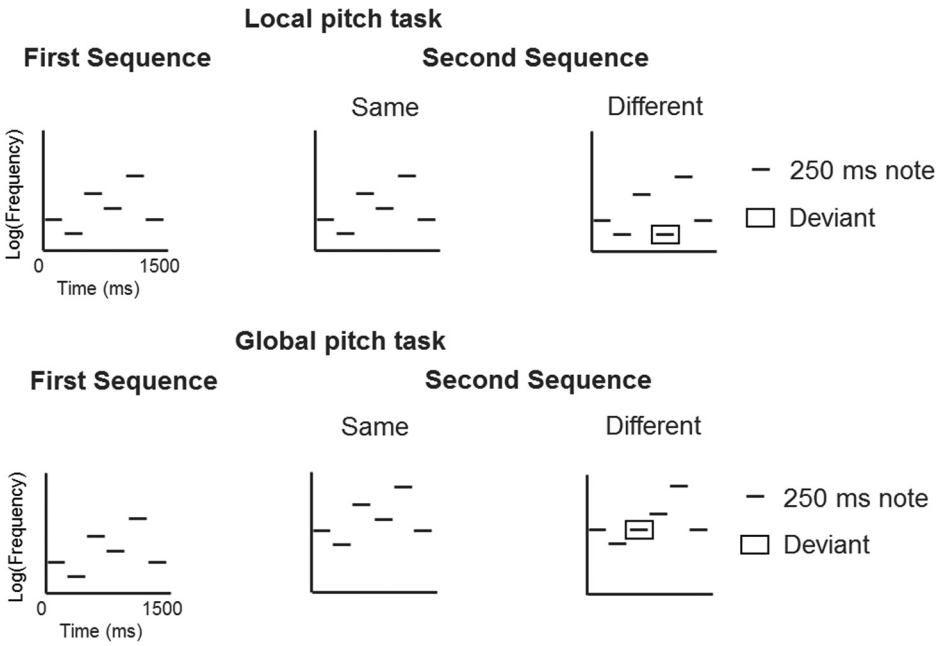

FIG. 2. Illustration of the local and global pitch perception tasks.

sequences were exactly the same. In the global pitch task, participants were instructed to pay attention to the overall pitch pattern, such that if the sequence changed in absolute values but kept the overall contour, they should respond with "s." The task was blocked by local versus global pitch perception with the order of task condition counterbalanced across listeners and all pairs of sequences in each pitch perception task were randomly played. Four practice trials were used before the test trials to ensure that participants understood the procedure of each task.

B. Results

1. Talker identification

The results are shown in Fig. 3. The dependent measure was the percentage of correct identifications in the generalization phase. We compared the groups' accuracy using a 3 between-subject (musical training: Extensive Training, Minimal Training and Non-Musicians) × 2 between-subject (native language: English and Mandarin) × 2 within-subject (language condition: English and Mandarin) ANOVA. The ANOVA revealed a significant main effect of musical training, $F(2, 128) = 5.939$, $p = 0.003$, $\eta_p^2 = 0.085$. *Post hoc* comparisons indicated a significant contrast between the Extensive Training group ($M = 0.65$, SD = 0.12) and Non-Musicians ($M = 0.56$, SD = 0.12), $p = 0.005$, a marginally significant difference between the Minimal Training group ($M = 0.62$, SD = 0.12) and Non-Musicians, $p = 0.061$, but no significant difference between the Extensive Training and Minimal Training groups. There was a significant main effect of native language, $F(1, 128) = 4.969$, $p = 0.028$, $\eta_p^2 = 0.037$ with Mandarin listeners ($M = 0.64$, SD = 0.13) outperforming English listeners ($M = 0.59$, SD = 0.12) in general. There was no main effect of language condition. There was an interaction between native language and language condition, reflecting the language familiarity effect, $F(1, 128) = 182.405$, $p < 0.001$, $\eta_p^2 = 0.588$. No other interaction effects were significant.

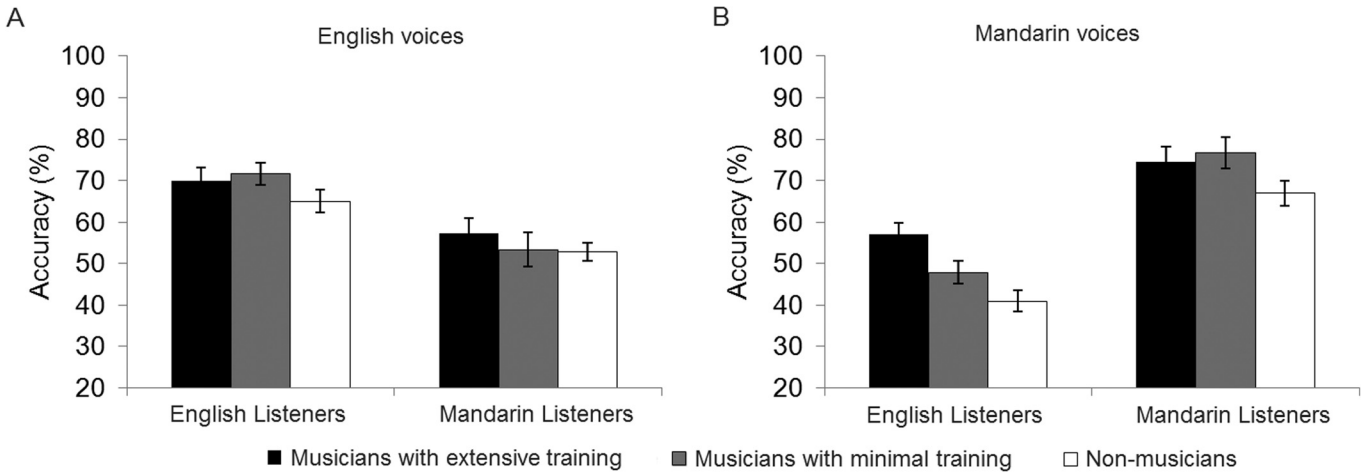

FIG. 3. Talker identification accuracy as a function of listener group for (A) English voices and (B) Mandarin voices (experiment 2). Error bars indicate ±1 SEM. All individuals scored above chance (20%), shown as baseline.

2. Pitch perception

Accuracy on each pitch task was calculated. Hit (H) and false alarm (FA) rates were further transformed into d′ scores (Fig. 4). In the cases where listeners had perfect accuracy (H = 1, FA = 0), a correction (H = 0.95, FA = 0.05) was made to avoid an infinite d′. Based on initial diagnostics and the Box–Cox procedure (Box and Cox, 1964), d′ scores were log-transformed to improve normality and homogeneity of variance assumptions necessary for a parametric ANOVA. Ten English participants (2 Extensive Training, 1 Minimal Training, and 7 Non-Musicians) and two Mandarin Non-Musicians had negative d′ scores (indicating below-chance performance) and were excluded for further analysis. Log-transformed d′ scores were then submitted to a repeated-measures ANOVA with music training (Extensive Training, Minimal Training, and Non-Musicians) and native language (English and Mandarin participants) as the between-subjects factors and task type (global and local) as the within-subject factor. There was a significant effect of task type, $F(1, 116) = 96.929$, $p < 0.001$, $\eta_p^2 = 0.455$, with the global task being generally harder than the local task; a significant effect of musical training, $F(2, 116) = 7.653$, $p < 0.001$, $\eta_p^2 = 0.117$, with *post hoc* comparisons revealing a significant contrast only between Extensive Training participants and Non-Musicians, $p < 0.001$; a significant effect of native language, $F(1, 116) = 46.318$, $p < 0.001$, $\eta_p^2 = 0.285$, with Mandarin listeners demonstrating higher accuracy across pitch tasks relative to native English listeners. Importantly, the overall findings regarding musical training and tone language experience paralleled the results on talker identification across listener groups. There was also an interaction between native language and task type, $F(1, 116) = 13.003$, $p < 0.001$, $\eta_p^2 = 0.101$. To unpack this interaction, we examined the two tasks separately to see if the observed benefits from tone language use are demonstrated differently in the two tasks. Results were collapsed across musical training given no interaction between this factor and task type. Given the above main effect of task type and native language group, we examined whether the task type effect was more pronounced in the English group than in the Mandarin group or vice versa. We calculated the difference between the performance for the local task and that for the global task. A *t*-test revealed that English participants exhibited larger difference between the global and the local pitch task than Mandarin participants, $t(120) = 3.784$, $p < 0.001$; this differential pattern was driving the interaction between task type and native language group.

3. The mediating effect of pitch perception on talker identification

Experiment 2 involved groups of participants that crossed within two profiles: musical training background (Extensive Training, Minimal Training, and Non-Musicians) and tone language experience (Mandarin speaking versus English-speaking). So far both predictions on the group level were confirmed: Experiment 2 replicated experiment 1 by showing more accurate voice identification among musicians relative to non-musicians and extended the results by showing a similar perceptual benefit among tone language users relative to speakers of an atonal language. Putting together the parallel findings in these two populations' enhanced pitch perception and the prominence of pitch in voice signatures, we hypothesized that individual pitch processing ability is the mediator that drives the greater voice perception accuracy in these two populations. Mediation analysis is a helpful tool to investigate potential causal hypotheses (Baron and Kenny, 1986) and has been widely utilized in psychological studies. To test the mediating effect of pitch processing ability on the relationship between musical/tone language experience and talker identification, we examined two independent variables that predicted pitch performance: musical experience (the number of years of musical training) as a continuous variable and native language (0 = English, 1 = Mandarin) as a dichotomous variable. Due to the language familiarity effect, the native language condition is much easier than the non-native language condition as we saw in the ANOVA analysis. Thus here we ran two separate mediation models, one for the non-native language condition and one for the native language condition. In both models, the dependent variable was the talker identification accuracy. The hypothesized mediator was as a composite measure: Average sensitivity score (log-transformed d′) across the two

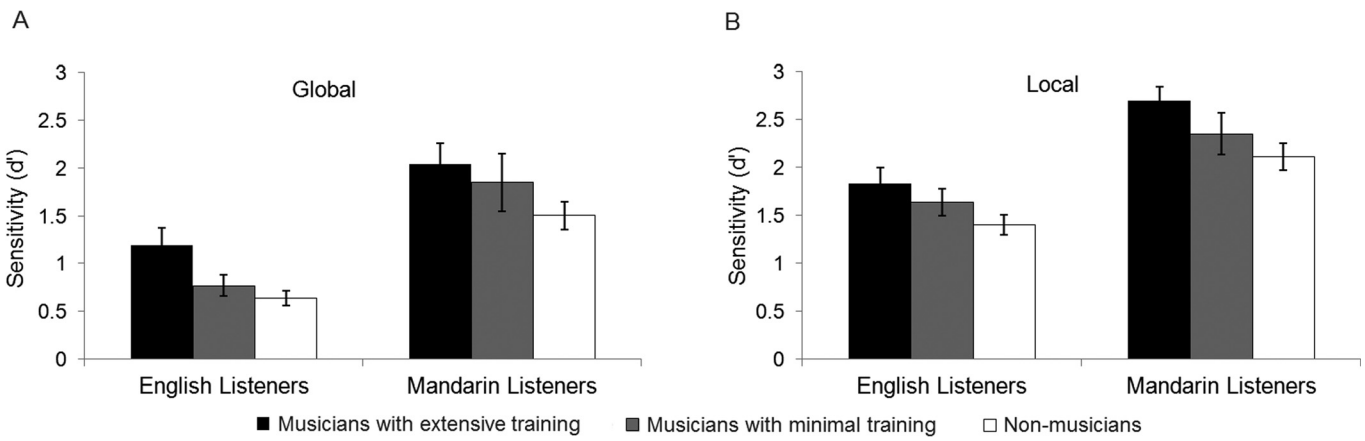

FIG. 4. Group comparisons of discrimination sensitivity in the global and local pitch perception tasks. Error bars represent ±1 SEM.

TABLE I. Correlations between talker identification performance (native and non-native language condition) and pitch sensitivity ($n = 122$) in experiment 2.

	Native	Non-native	Global
Non-native	0.370[a]		
Global	0.214[b]	0.334[a]	
Local	0.254[c]	0.258[c]	0.580[b]

The significance level reflects an uncorrected alpha criterion.
[a]$p < 0.001$.
[b]$p < 0.05$.
[c]$p < 0.01$ (2-tailed).

pitch tasks. This composite measure was used because a preliminary analysis indicated high within-subject correlation between local and global pitch perception. This result is shown in Table I. For this reason, we did not consider local and global sensitivity as separate mediators. A mediation hypothesis is usually represented in a path diagram. In the diagram, path coefficients denote the strength of hypothesized causal relations. Regression coefficient c represents the total effect of an initial independent variable (IV) on the dependent variable (DV). Coefficient a represents the influence of the IV on the mediating variable; b represents the influence of mediator on the DV; the product a × b, or ab estimates the strength of the indirect effect of IV on DV, that is, how much of the increase in the DV that occurs as the IV increases is due to changes in the mediator; c′ denotes the strength of the direct effect of the IV on the DV, that is, any effect that is not mediated by the mediating variable. The total effect c is the sum of the direct effect of the IV on the DV (c') and the indirect effect (ab) of IV on the DV through the mediator. We implemented the mediation analyses using the bootstrapping method (Preacher and Hayes, 2004). Refer to Fig. 5 for the path diagram with standardized path coefficients.

a. Non-native language condition. The total effect of musical experience on talker identification was significant, $c = 0.238$, $t(119) = 2.699$, $p = 0.008$; musical experience was significantly predictive of the hypothesized mediator, pitch, $a = 0.317$, $t(119) = 4.266$, $p < 0.001$; pitch was significantly predictive of talker identification, $b = 0.264$, $t(118) = 2.478$, $p = 0.015$. The estimated direct effect of musical training on talker identification, controlling for pitch was nonsignificant, $c' = 0.154$, $t(118) = 1.665$, $p = 0.10$. Furthermore, the indirect effect of musical experience on talker identification, $ab = 0.083$, was significant, with a 95% bias-corrected and accelerated bootstrap confidence interval of [0.001, 0.007] which did not include zero (Preacher and Hayes, 2004).

The total effect of native language (tone language or non-tone language) on talker identification was significant, $c = 0.221$, $t(119) = 2.510$, $p = 0.013$; native language was significantly predictive of the hypothesized mediator, pitch, $a = 0.545$, $t(119) = 7.344$, $p < 0.001$. The estimated direct effect of native language on talker identification, controlling for pitch became nonsignificant, $c' = 0.077$, $t(118) = 0.743$, $p = 0.46$. Furthermore, the indirect effect of native language

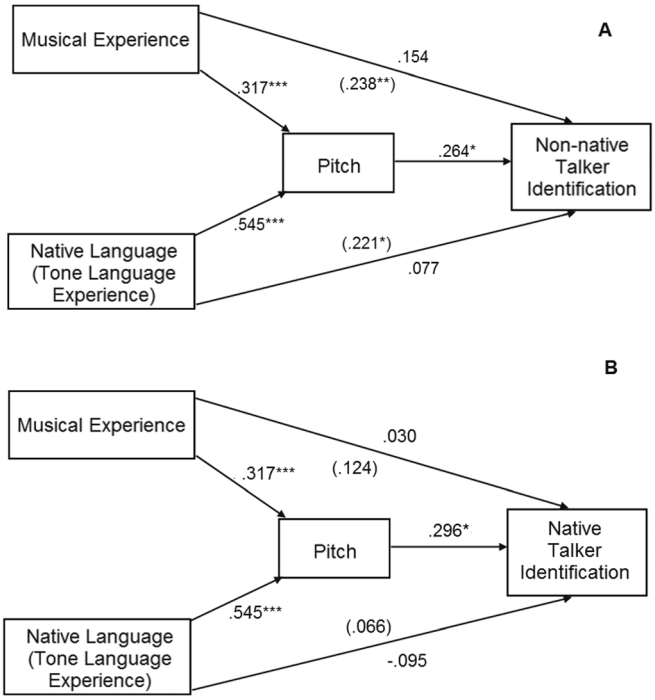

FIG. 5. (A) The path diagram shows pitch as a mediator of the link between musical/tone language experience and talker identification in non-native languages. (B) The path diagram shows no direct or indirect influence of musical/tone language experience on talker identification in the native language condition. All path coefficients are standardized and asterisks indicate significant coefficients (*$p <$. 05; **$p < 0.01$; ***$p < 0.001$). Coefficients in parentheses indicate relationship before controlling for the mediating factor of pitch.

on talker identification, $ab = 0.144$, was significant, as suggested by a 95% bias-corrected and accelerated bootstrap confidence interval of [0.008, 0.084]. Thus as hypothesized, individual pitch perception mediated the influence of musical and linguistic experience on talker identification.

b. Native language condition. The total effect of musical training on talker identification was nonsignificant, $c = 0.124$, $t(119) = 1.36$, $p = 0.18$. The total effect of native language on talker identification was nonsignificant, $c = 0.066$, $t(119) = 0.723$, $p = 0.47$. Thus no influence of either musical or linguistic experience was observed for talker identification in the native language condition.

C. Discussion

The results successfully replicated and extended our findings in experiment 1: Again as a group, musicians were better at talker identification than non-musicians. As hypothesized, this influence is (at least in part) due to enhanced pitch processing abilities. First, the amount of musical experience positively predicted performance in the pitch tasks. Furthermore in the mediation analysis, the significant effect of musical training on talker identification (significant only in the non-native language condition) became nonsignificant after controlling the effect of pitch perception.

The contrast between listeners who speak a tone language versus a non-tone language was also revealing: Relative to English listeners matched on musical training, Mandarin listeners exhibited significantly higher performance in both the pitch tasks and talker identification in general. Again, pitch perception ability mediated the relationship between linguistic experience with tones and talker identification performance.

One important thing to note is that as clearly demonstrated by the differential results of the mediation analyses, the facilitatory effect of pitch sensitivity exists only in the non-native language condition, replicating the finding from experiment 1. Linguistic information (e.g., phonological, lexical, syntactic) specific to one's native language is absent in unfamiliar language conditions. Under these circumstances, we suggest that the ability to identify talkers relies more heavily on non-linguistic information such as listeners' pitch processing abilities. In sum, the results provided solid empirical evidence that individual differences in pitch processing ability exist and that they account for the individual variability observed in talker identification.

IV. GENERAL DISCUSSION

In this study, we sought to characterize individual differences in how well typical adults can perceive pitch and use it to identify talkers. We hypothesized that increased sensitivity to pitch would increase accuracy in talker identification, a task that is cognitively much more difficult than basic pitch perception. We addressed this question on two levels: First, we asked whether musicians and tone language users, two populations who have been shown to possess enhanced pitch processing, also have enhanced perception of talker identity; second, we asked whether an individual's pitch perception ability is directly related to his or her ability to identify talkers in general. Furthermore, we assessed whether pitch expertise acquired through musical practice or years of tone language use accounts for perceptual benefits in musicians and tone language users.

A. The impact of musical training on pitch and talker identification

Long-term musical training is known to sharpen pitch acuity (Bidelman *et al.*, 2013; Tervaniemi *et al.*, 2005). The perceptual benefit of musical training is not limited to musical pitch but extends to linguistic use of pitch. In particular, a strong link between music and language emerges in studies assessing processing of lexical tones. Musical training/aptitude predicts non-tone language speakers' performance in lexical tone identification (Delogu *et al.*, 2010) and imitation (Gottfried *et al.*, 2004) and their ability to learn lexical tones in a lexical identification task (Wong and Perrachione, 2007).

In the current study, amateur musicians demonstrated superiority in detecting pitch changes (both global and local) in pure tones compared to musically naive individuals. Furthermore, our finding expands previous research of musical influence on musical and linguistic pitch processing to a new area: Voice identity processing. We confirmed in two experiments that musical training does have a role in promoting more accurate talker identification. Unlike perception of musical notes/melodies or lexical tones in tone languages, the use of pitch in talker identification is not explicit. In the talker identification task, listeners presumably can make use of a combination of various perceptual cues. And indeed, listeners rely heavily on linguistic cues when perceiving talkers in one's native language as demonstrated by the language familiarity effect in the current study and in previous literature (e.g., Perrachione *et al.*, 2009). However, when an unfamiliar language is spoken, we observed a clear benefit of musicianship on voice perception in both experiments without explicitly directing listeners' attention to pitch variation. Moreover, the number of years of musical training positively contributed to accuracy in talker identification, demonstrating a cross-domain impact of music experience. The mediation analysis revealed that this impact is achieved via enhanced pitch processing. In particular, it indicated that processing talker identity depends, at least partially, on domain-general processes shared with music and thus can be sharpened by long-term musical training.

It is also important to note, however, that although pitch processing mediates the link between musical training and talker identification, our analysis does not rule out the possibility that other perceptual or cognitive variables may also be at play. On top of enhanced processing pertaining to pitch, musicians are found to be more sensitive to timbre than non-musicians (Chartrand and Belin, 2006), an acoustic property assumed to contribute to voice quality and emotion in speech (Juslin and Laukka, 2003). Meanwhile, musicians are also reported to outperform non-musicians in other language tasks such as second language production and perception, pitch memory, verbal memory, and perhaps segmental processing (Bidelman *et al.*, 2013; Chan *et al.*, 1998; Slevc and Miyake, 2006; Marie *et al.*, 2011; although see Delogu *et al.*, 2010). In the current study, the talker identification task in each language condition involved five different talkers and was cognitively more complex than pitch perception. Thus it is possible that musicians' advantages in timbre processing, memory capacity, or other cognitive factors also contribute to the superiority in talker identification performance. Moreover, it is possible that pre-existing capacities for auditory processing in populations who pursue musical training are really the root of better performance in talker identification rather than experience-related gains in pitch processing over the course of musical training. However, the close relationship between years of musical study and talker identification skills observed in the current study questions this interpretation and points to a training-related benefit in musicians. Further study is needed to see if other cognitive benefits of musicianship not originating from pitch processing share the contribution to enhanced voice perception.

B. The impact of tone language experience on pitch and talker identification

Native speakers of tone languages are found to possess enhancements in pitch-related abilities due to long-term use of tones in language. These advantages include enhanced

perceptual sensitivity to non-linguistic uses of pitch (Bidelman *et al.,* 2013; Bradley, 2012; Pfordresher and Brown, 2009) as well as enhanced or more precise neural responsiveness to pitch changes (Krishnan, *et al.,* 2009; Chandrasekaran *et al*., 2007).

The current study assessed Mandarin listeners and English listeners' perception of pure tones, measuring their sensitivity to changes between two tone sequences: In pitch height or in overall contour. Previous neurophysiological evidence suggests that musicians can exploit both interval-based pitch cues (Tervaniemi *et al.,* 2005), similar to those assessed by the local pitch task, and contour-based cues (Wayland *et al.,* 2010), similar to the type of stimuli used in the global task. In contrast, Bradley (2012) found that a benefit from lexical tone use only when perceiving pitch contours but not local changes in pitch intervals (but see Bidelman *et al.,* 2013). We thus expected to observe different patterns in musicians and tone language speakers in their performance on the local and global pitch tasks. However, Mandarin listeners demonstrated substantial advantages over English counterparts consistently in both tasks, controlling for musical training. In addition, our attempt to use this dissociation to assess the relative contribution of local pitch and global pitch skills in promoting talker identification was inconclusive due to the high correlation between the two measures. Nevertheless, an important new finding of the current study is that for the first time, we demonstrated that highly similar to musicians, tone language speakers' perceptual enhancement is not restricted to cognitive tasks that require explicit attention to pitch but also clearly generalizes to the ability to perceive and identify talkers in a non-native context.

It is worth noting that the bilingual experience of our Mandarin listeners could be a possible alternative interpretation to account for the perceptual benefits. For instance, L2 language ability has found to have gradient effect on talker identification (Bregman and Creel, 2014). On the other hand, reports on the auditory processing advantages originating from bilingualism have only been reported in studies testing high-proficient early bilinguals (Krizman *et al.,* 2012). It is an empirical question whether similar benefits are present at all in late bilinguals, such as the Mandarin listeners in the current study. Finally, while bilingualism may play a role in the superior performance of the Mandarin group, our mediation analysis in experiment 2 clearly demonstrated that pitch processing abilities explained a large proportion of the effect of native language experience on talker identification in the non-native listening condition.

C. Shared mechanisms: Music, language and voice perception

The convergent results on enhanced talker identification as a function of experience in the music domain and the language domain have important implications regarding the plasticity of auditory perception. Such a finding suggests that the functional or structural changes effected by either musical exercise or long-term lexical tone use heighten listeners' sensitivity to the pitch dimension in a general way. The results speak directly against strict modularity of cognitive systems involved in music, language, and, further, talker identity.

As mentioned previously, there is already ample evidence showing that experience-dependent pitch expertise acquired through long-term exposure to tones in the speech input produces changes in music perception and vice versa. Such bidirectional transfer between the music and language domains has been taken as evidence supporting a shared auditory processing system through which domain-specific experience attunes domain-general auditory skills (Kraus and Chandrasekaran, 2010; Skoe and Kraus, 2012). Our study reveals that on top of between-domain transfers, the enhanced pitch processing percolates to human voice perception. It is important to consider the differences between talker identification task and previous tasks pertaining to pitch use. Putting pitch use in the musical domain aside, in speech, pitch can be used linguistically to contrast word meanings or sentence type, or used paralinguistically to denote speakers' emotions or identity. It is well documented that musicians excel at detecting subtle pitch changes in lexical tones (Delogu *et al.,* 2010) and are sensitive to pitch contours that differentiate a statement from a question (e.g., Deguchi *et al.,* 2012) as do tone language speakers (e.g., Stevens *et al.,* 2013). Moreover, suggestive evidence exists showing a relationship between individual differences in pitch processing on speech and non-speech tasks, indicating that the use of pitch in linguistic and non-linguistic domains relies on overlapping resources (Perrachione *et al.,* 2013). However, all these tasks are designed such that listeners' attention is directed toward the pitch patterns by the task requirements. One exception is provided by studies that show musicians' better encoding of speakers' emotions than non-musicians (Thompson *et al.,* 2004). In the current study, the nature of talker identification tasks leaves it to a listener's own discretion to selectively make use of salient information in the sound signal. For one's native language, various cues (phonological, lexical, and syntactic) other than pitch can more readily encode a speaker's characteristics. In the non-native language condition, the linguistic information was stripped away and pitch stood out as a crucial information-bearing dimension. We speculate that the difference in the availability of various sources of information led to the clear contrast between native and non-native language conditions. Across two experiments, no differences were observed in the native language condition for either of the expert groups. Given the difficulty of the talker identification task, listeners' performance was far from ceiling even in the native conditions; yet the musical training or tone experience did not produce additional advantages in making use of linguistic elements to decipher talker identity. In contrast, the effective representation of pitch in musicians and tone language users made them the better performers in the non-native language condition. Such results align well with the theoretical framework of the OPERA hypothesis (Patel, 2011, 2012). The hypothesis provides a clear and specific rationale for the reason why musical training can benefit neural encoding of speech processing networks. Namely, higher demands on the accuracy of pitch processing in musical

training drive the auditory system to function with higher precision than usually required by ordinary speech processing tasks. A critical assumption of the theory is that the auditory system tolerates "good-enough" processing in most circumstances; and individuals can vary substantially in pitch-processing precision just to meet the needs of typical speech communication. When it comes to musical performance, higher precision is needed; and thus the auditory system adapts in the face of such demand. Similarly, we saw in the current study that when redundant cues are available to perceive talker identity in the native language condition, musical or tone language experience does not confer further benefit behaviorally. The OPERA hypothesis can be extended to account for the experience-dependent plasticity elicited by tone language experience as well. In both musical training and long-term use of tone languages, the pitch processing system is retuned via prolonged learning of these cues and leads to enhanced sensitivity to pitch, which prepares listeners for good performance. Long-term explicit training or exposure to pitch use in musicians and tone language speakers reshapes the perceptual system and makes it more adaptive in general contexts. Furthermore, consistent with the OPERA hypothesis, the domain-general to task-specific transfer occurred only when the relevant sensory and cognitive processes became a bottleneck to performance as observed in the non-native language condition. Exactly in and only in this condition, perceptual experts outperform untrained individuals in using pitch to address talker identity.

V. CONCLUSION

We observed a direct link between individual differences in pitch perception and talker identification. First, musicianship predicted enhanced pitch processing skills and enhanced ability to accurately identify unfamiliar talkers. In parallel, the facilitative effect of good pitch perception skills was similarly found in native tone language speakers. Second, individual pitch processing ability mediated the impact of musical and linguistic influence on talker identification. Theoretically our results support a shared resources hypothesis regarding music, language, and voice perception, suggesting that skills obtained via domain-specific training can be transferred to processing of talker identity embedded in speech sounds. This finding adds a new dimension to the existing literature that demonstrates transfer effects between music perception and speech perception.

ACKNOWLEDGMENTS

We are grateful to Carol Fowler for her advice throughout the course of this project. We also thank David Kenny for generous and valuable help on the mediation analysis. We acknowledge support from "Fund for Innovation Fund in Science Education" at University of Connecticut. This work was supported by NIH R03 DC009495 (E. Myers, PI) and by NIH P30 DC010751 (D. Lillo-Martin, PI). The content is the responsibility of the authors and does not necessarily represent official views of the NIH or NIDCD.

APPENDIX: STIMULUS SENTENCES IN TALKER IDENTIFICATION TASK

1. Mandarin sentences

他到过很多地方观光旅游
ta dao guo hen duo di fang guan guang lv you
"He has visited a lot of places."
杜鹃的叫声报告了春天的来临
du juan de jiao sheng bao gao le chun tian de lai lin
"The cuckoo reports the coming of spring."
马上就要转播棒球比赛了
ma shang jiu yao zhuan bo bang qiu bi sai le
"The baseball game will be on the air in a few seconds."
夏日大平原的夕阳尤其美丽
xia ri da ping yuan de xi yang you qi mei li
"The sunset on the prairie is especially beautiful in the summer."
下雪以后,田野里白皑皑的一片
xia xue yi hou, tian ye li bai ai ai de yi pian
"After the snow, the field was a snow-white patch."
院子门口不远处就是一个地铁站
yuan zi men kou bu yuan chu jiu shi yi ge di tie zhan
"There is a subway station not far from the entrance to the yard."
外宾们十分喜爱湖上的景色
wai bin men shi fen xi ai hu shang de jing se
"The foreign guests reveled in the scenery of the lake."
这是一个休息散心的好去处
zhe shi yi ge xiu xian san xin de hao qu chu
"This is a good place to go to relax."
山间的小道蜿蜒曲折
shan jian de xiao dao wan yan qu zhe
"The mountain path winds torturously."
那小女孩学着梳理羊毛
na xiao nv hai xue zhe shu li yang mao
"The little girl was learning to comb wool."

2. Spanish sentences

Su manera de recitar le encantó al público
"The audience was enchanted by his recital."
Le regaló una piñata de cumpleaños
"She gave him a birthday piñata."
Los señores están en buena posición económica
"The gentlemen are well off."
Notifiqué a todos los miembros
"Notify all the members."
Tiene la suerte de vivir en una casa grande
"She is lucky enough to live in a big house."
Es una persona de conversación amena
"She is always very nice to talk to."
Debo entrar a comprar cigarillos
"I need to go in to buy some cigarettes."
A la gente le gusta la música que es bonita
"People enjoy beautiful music."
La próxima Navidad la pasaré con mi gente
"I am spending next Christmas with my family."
Estuvimos de vacaciones en Sudamérica
"We spent our holidays in South America."

3. English sentences

Her yellow purse was full of useless trash.
She saw a cat in the neighbor's apartment.
Victoria has a great variety of candies.
Try angling the camera for a more interesting picture.
The girl wiped the grease off his dirty face.
The oldest birch looked stark white and lonesome.
Nobody has found the silver necklace with hearts on it.
All the teams have to compete for the championship team.
He was on the verge of telling me all the secret.
We do not think the senator should run for reelection.

[1]All participants completed an additional sentence-in-noise recognition task after the completion of the talker identification task. This task was used as an assessment for other purposes. Given the focus of the current paper, it was not discussed here.
[2]All *post hoc* comparisons were conducted with a Bonferroni correction.
[3]We maintained a categorical grouping of musical factors to make our results comparable to previous studies (e.g., Bidelman *et al.*, 2013) on the benefit of musical experience, which generally contrasted musicians with non-musicians.
[4]A short phonetic categorization task was completed by all participants for other purposes. Preliminary analysis revealed that performance on this task was not related to pitch perception performance. The results were not discussed here given the focus of this paper.

Baron, R. M., and Kenny, D. A. (**1986**). "The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations," J. Personality Social Psychol. **51**, 1173–1182.
Baumann, O., and Belin, P. (**2010**). "Perceptual scaling of voice identity: Common dimensions for different vowels and speakers," Psychol. Res. **74**(1), 110–120.
Belin, P., Bestelmeyer, P. G., Latinus, M., and Watson, R. (**2011**). "Understanding voice perception," Br. J. Psychol. **102**(4), 711–725.
Bidelman, G. M., Gandour, J. T., and Krishnan, A. (**2011**). "Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem," J. Cognit. Neurosci. **23**(2), 425–434.
Bidelman, G. M., Hutka, S., and Moreno, S. (**2013**). "Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: Evidence for bidirectionality between the domains of language and music," PLoS One **8**(4), e60676.
Box, G. E. P., and Cox, D. R. (**1964**). "An analysis of transformations," J. R. Stat. Soc. Ser. B Methodol. **26**(2), 211–252.
Bradley, E. D. (**2012**). "Tone language experience enhances sensitivity to melodic contour," in *Linguistic Society of America Annual Meeting Extended Abstracts*, January 5-8, Portland, OR.
Bregman, M. R., and Creel, S. C. (**2014**). "Gradient language dominance affects talker learning," Cognition **130**(1), 85–95.
Burnham, D., Brooker, R., and Reid, A. (**2014**). "The effects of absolute pitch ability and musical training on lexical tone perception," Psychol. Music (in press).
Chan, A. S., Ho, Y., and Cheung, M. (**1998**). "Music training improves verbal memory," Nature **396**, 128.
Chandrasekaran, B., Krishnan, A., and Gandour, J. T. (**2007**). "Experience-dependent neural plasticity is sensitive to shape of pitch contours," Neuroreport **18**(18), 1963–1967.
Chartrand, J. P., and Belin, P. (**2006**). "Superior voice timbre processing in musicians," Neurosci. Lett. **405**(3), 164–167.
Cleary, M., and Pisoni, D. B. (**2002**). "Talker discrimination by prelingually deaf children with cochlear implants: Preliminary results," Ann. Otol. Rhinol. Laryngol. Supple. **189**, 113–118.
Creel, S. C., and Bregman, M. R. (**2011**). "How talker identity relates to language processing," Lang. Linguist. Compass **5**(5), 190–204.
Deguchi, C., Boureux, M., Sarlo, M., Besson, M., Grassi, M., Schön, D., and Colombo, L. (**2012**). "Sentence pitch change detection in the native and unfamiliar language in musicians and non-musicians: Behavioral, electrophysiological and psychoacoustic study," Brain Res. **1455**, 75–89.
Delogu, F., Lampis, G., and Belardinelli, M. O. (**2010**). "From melody to lexical tone: Musical ability enhances specific aspects of foreign language perception," Eur. J. Cognit. Psychol. **22**(1), 46–61.
Foxton, J. M., Talcott, J. B., Witton, C., Brace, H., McIntyre, F., and Griffiths, T. D. (**2003**). "Reading skills are related to global, but not local, acoustic pattern perception," Nat. Neurosci. **6**(4), 343–344.
Francis, A. L., and Driscoll, C. (**2006**). "Training to use voice onset time as a cue to talker identification induces a left-ear/right-hemisphere processing advantage," Brain Lang. **98**(3), 310–318.
Gaab, N., and Schlaug, G. (**2003**). "The effect of musicianship on pitch memory in performance matched groups," Neuroreport **14**(18), 2291–2295.
Gelfer, M. P. (**1988**). "Perceptual attributes of voice: Development and use of rating scales," J. Voice **2**, 320–326.
Gottfried, T. L., Staby, A. M., and Ziemer, C. J. (**2004**). "Musical experience and Mandarin tone discrimination and imitation," J. Acoust. Soc. Am. **115**, 2545.
Juslin, P. N., and Laukka, P. (**2003**). "Communication of emotions in vocal expression and music performance: Different channels, same code?," Psychol. Bull. **129**(5), 770–814.
Kraus, N., and Chandrasekaran, B. (**2010**). "Music training for the development of auditory skills," Nat. Rev. Neurosci. **11**(8), 599–605.
Krishnan, A., Swaminathan, J., and Gandour, J. T. (**2009**). "Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context," J. Cognit. Neurosci. **21**(6), 1092–1105.
Krizman, J., Marian, V., Shook, A., Skoe, E., and Kraus, N. (**2012**). "Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages," Proc. Natl. Acad. Sci. U.S.A. **109**(20), 7877–7881.
Künzel, H. J. (**1994**). "On the problem of speaker identification by victims and witnesses," Forensic Linguist. **1**(1), 45–57.
Magnuson, J. S., and Nusbaum, H. C. (**2007**). "Acoustic differences, listener expectations, and the perceptual accommodation of talker variability," J. Exp. Psychol. Hum. Percept. Perform. **33**, 391–409.
Marie, C., Delogu, F., Lampis, G., Belardinelli, M. O., and Besson, M. (**2011**). "Influence of musical expertise on segmental and tonal processing in Mandarin Chinese," J. Cognit. Neurosci. **23**(10), 2701–2715.
Mary, L., and Yegnanarayana, B. B. (**2008**). "Extraction and representation of prosodic features for language and speaker recognition," Speech Comm. **50**(10), 782–796.
Patel, A. D. (**2011**). "Why would musical training benefit the neural encoding of speech? The OPERA hypothesis," Front. Psychol. **2**, 142.
Patel, A. D. (**2012**). "The OPERA hypothesis: Assumptions and clarifications," Ann. N.Y. Acad. Sci. **1252**(1), 124–128.
Peretz, I. (**1990**). "Processing of local and global musical information by unilateral brain-damaged patients," Brain **113**, 1185–1205.
Perrachione, T. K., Del Tufo, S. N., and Gabrieli, J. D. (**2011**). "Human voice recognition depends on language ability," Science **333**(6042), 595.
Perrachione, T. K., Fedorenko, E. G., Vinke, L., Gibson, E., and Dilley, L. C. (**2013**). "Evidence for shared cognitive processing of pitch in language and music," PLoS One **8**(8), e73372.
Perrachione, T. K., Pierrehumbert, J. B., and Wong, P. (**2009**). "Differential neural contributions to native- and foreign-language talker identification," J. Exp. Psychol. Hum. Percept. Perform. **35**(6), 1950–1960.
Perrachione, T. K., and Wong, P. (**2007**). "Learning to recognize speakers of a non-native language: Implications for the functional organization of human auditory cortex," Neuropsychologia **45**(8), 1899–1910.
Pfordresher, P. Q., and Brown, S. (**2009**). "Enhanced production and perception of musical pitch in tone language speakers," Atten. Percept. Psychophys. **71**(6), 1385–1398.
Preacher, K. J., and Hayes, A. F. (**2004**). "SPSS and SAS procedures for estimating indirect effects in simple mediation models," Behav. Res. Methods Instrum. Comput. **36**, 717–731.
Remez, R. E., Fellowes, J. M., and Rubin, P. E. (**1997**). "Talker identification based on phonetic information," J. Exp. Psychol. Hum. Percept. Perform. **23**, 651–666.
Sanders, L. D., and Poeppel, D. (**2007**). "Local and global auditory processing: Behavioral and ERP evidence," Neuropsychologia **45**, 1172–1186.
Schmidt-Nielsen, A., and Crystal, T. (**1998**). "Human vs. machine speaker identification with telephone speech," in *Proceedings of ICSLP 98*.
Skoe, E., and Kraus, N. (Editors) (**2012**). "Human subcortical auditory function provides a new conceptual frame work for considering modularity," in *Language and Music as Cognitive Systems* (Oxford University Press, Oxford, UK), pp. 269–282.
Slevc, L. R., and Miyake, A. (**2006**). "Individual differences in second-language proficiency does musical ability matter?," Psychol. Sci. **17**(8), 675–681.

Strait, D. L., Kraus, N., Parbery-Clark, A., and Ashley, R. (**2010**). "Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance," Hear. Res. **261**, 22–29.
Stevens, C. J., Keller, P. E., and Tyler, M. D. (**2013**). "Tonal language background and detecting pitch contour in spoken and musical items," Psychol. Music **41**(1), 59–74.
Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., and Schröger, E. (**2005**). "Pitch discrimination accuracy in musicians versus nonmusicians: An event-related potential and behavioral study," Exp. Brain Res. **161**(1), 1–10.
Thompson, W. F., Schellenberg, E. G., and Husain, G. (**2004**). "Decoding speech prosody: Do music lessons help?," Emotion **4**(1), 46–64.
Wayland, R., Herrera, E., and Kaan, E. (**2010**). "Effects of musical experience and training on pitch contour perception," J. Phonet. **38**(4), 654–662.
Winters, S. J., Levi, S. V., and Pisoni, D. B. (**2008**). "Identification and discrimination of bilingual talkers across languages," J. Acoust. Soc. Am. **123**, 4524–4538.
Wong, P. C. M., and Perrachione, T. K. (**2007**). "Learning pitch patterns in lexical identification by native English-speaking adults," Appl. Psycholinguist. **28**, 565–585.