LANGUAGE
TEACHING
RESEARCH

Article

Language Teaching Research
2023, Vol. 27(4) 1016–1031

DOI: 10.1177/1362168820971791
journals.sagepub.com/home/ltr
SAGE

Harnessing the musician advantage: Short-term musical training affects non-native cue weighting of linguistic pitch

Seth Wiener
Carnegie Mellon University, USA

Evan D. Bradley
Penn State Brandywine, USA

Abstract

Lexical tone languages like Mandarin Chinese require listeners to discriminate among different pitch patterns. A syllable spoken with a rising pitch (e.g. *bí* 'nose') carries a different meaning than the same syllable spoken with a falling pitch (e.g. *bì* 'arm'). For native speakers (L1) of a non-tonal language, accurate perception of tones in a second language (L2) is notoriously difficult. Musicians, however, have typically shown an aptitude for lexical tone learning due to the unique perceptual demands of music. This study tested whether musical effects can be exploited to improve linguistic abilities in the general population. A pre-test, 8-week training, post-test design was used to measure L1 English participants' sensitivity to tone. Individual Differences Scaling was used to measure participants' weighting of pitch height and movement cues. Participants took part in classroom Mandarin learning only (+L2), musical ear training only (+Music), or classroom learning combined with musical training (+L2+Music). An L1 Mandarin group served as a baseline. At pre-test, mean sensitivity to tone and multidimensional scaling results were similar across all three L1 English groups. After training, all three L1 English groups improved in mean sensitivity, though only the +L2+Music group did so at a significant rate. Multidimensional scaling revealed that all groups increased their weighting of the more informative pitch movement cue at roughly equal rates. Short-term musical training thus affected change in cue weighting of linguistic pitch in a manner comparable to that occurring after a semester of L2 classroom learning. When combined with classroom learning, short-term musical training resulted in even greater sensitivity to pitch movement cues. These results contribute to models of music-language interaction and suggest that focused application of non-linguistic acoustic training can improve phonetic perception in ways that are relevant to language learning.

Corresponding author:
Seth Wiener, Department of Modern Languages, Carnegie Mellon University, 341 Posner Hall, 4980 Margaret Morrison, Pittsburgh, PA 15213, USA.
Email: sethw1@cmu.edu

Keywords
cue weighting, Mandarin lexical tone, music, pitch perception, second language learning

I Introduction

Accurate discrimination of speech sounds requires perception of multiple acoustic cues. For example, voice onset time (VOT) is a cue that represents the time between a stop consonant's release and the onset of voicing (see Abramson & Whalen, 2017 for a review). VOT is a useful cue for distinguishing an English /p/ (a 'voiceless' sound) from a /b/ (a 'voiced' sound) as in 'pat' and 'bat'. Formants are cues that represent the spectral shape or acoustic energy of speech. Formants are useful cues for distinguishing an English /ɹ/ (lower onset frequency of the third formant) from an /l/ (higher onset frequency of the third formant) as in 'rock' and 'lock' (Espy-Wilson, 1992). A speaker's fundamental frequency is a cue that represents vocal fold vibration. The psychological correlate of fundamental frequency is pitch. Pitch is a useful cue for distinguishing the Mandarin Chinese tone categories as in *yi1* (一 'one') from *yi4* (亿 'one million'; Ho, 1976; Howie, 1976).

These examples illustrate that acoustic cues do not contribute equally across speech sound categories or languages (see Schertz & Clare, 2020 for a review). The variability in how speech cues contribute to each category, and across languages, is commonly referred to as 'cue weighting' (Holt & Lotto, 2006; Toscano & McMurray, 2010; Toscano et al., 2010). Listeners appear to weight certain cues greater when they contribute to category membership, and less when they do not directly contribute to category membership (e.g. Escudero, Benders, & Lipski, 2009; Lipski, Escudero, & Benders, 2012). Accurate second language (L2) speech perception thus requires learning to weight various acoustic cues differently from how they are weighted in the first language (L1; Escudero & Boersma, 2004; Iverson et al., 2003).

In this study, we examine the cue weighting of pitch, the perceptual correlate of fundamental frequency (F0). Pitch is an important cue to Mandarin speech communication, and as we outline below, a difficult cue for L2 learners to acquire. Novel to the present study, we compare how different pitch learning modalities – linguistic, music, and their combination – affect how pitch height and movement cues are weighted by adult L2 Mandarin learners.

1 Cue weighting in an L2

Adults typically demonstrate poor cue weighting and limited categorical discrimination of novel speech sounds (e.g. Saito, 2015; Ylinen et al., 2010). Foundation evidence for poor L2 cue weighting comes primarily from adult L1 Japanese-L2 English learners. Japanese does not use /ɹ/ and /l/ to distinguish word meaning and instead possesses a single speech category that overlaps with English /ɹ/ and /l/ in acoustic space (Goto, 1971; Miyawaki et al., 1975). From a cue weighting perspective, L1 Japanese-L2 English learners struggle to perceptually weight third formant onset cues (e.g. Ingvalson, Holt, &

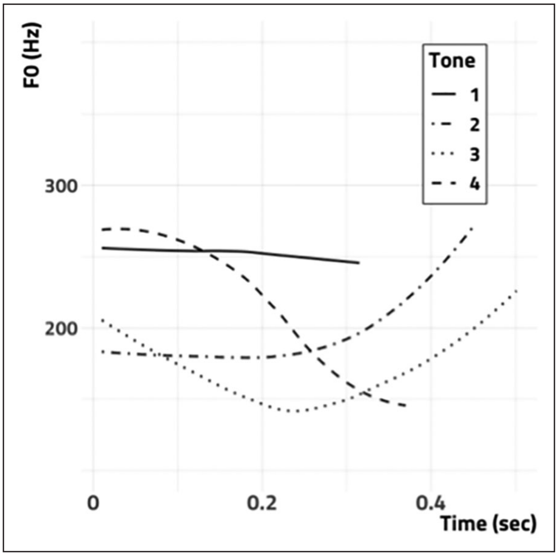

Figure 1. The four Mandarin tones spoken in isolation by a native speaker.

McClelland, 2012). As a result, discriminating and categorizing English /ɹ/ and /l/ categories can be incredibly difficult for L1 Japanese-L2 English learners, even after years of living in an English-speaking environment (e.g. Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Flege, Yeni-Komshian, & Liu, 1999; MacKain, Best, & Strange, 1981; Saito, 2013).

Another commonly tested L2 group – and the focus of the present study – is L1 English-L2 Mandarin learners (see Pelzl, 2019 for a review). L1 English and L1 Mandarin listeners weight a speaker's pitch cues differently (Chandrasekaran, Sampath, & Wong, 2010). This difference in cue weighting reflects the role pitch information plays in each language. In spoken English, pitch cues carry pragmatic, emotional, or indexical information (Gussenhoven, 2004). L1 English listeners therefore rely on pitch height cues as these cues can carry indexical and pragmatic information (Perrachione, Lee, Ha, & Wong, 2011; Wong & Perrachione, 2007).

In spoken Mandarin, however, pitch cues carry lexical information, i.e. tone. Figure 1 plots the four canonical tones spoken in isolation by a native female speaker from Beijing, the regional variety of speech many L2 Mandarin pedagogical materials emphasize. These tones can be summarized by their pitch patterns (Gandour, 1978; Ho, 1976). Tone 1 involves a relatively high-level pitch. Tone 2 involves a low to rising pitch. Tone 3 involves a low dipping pitch. Tone 4 involves a high to low falling pitch. L1 Mandarin listeners therefore rely on pitch movement across a syllable as this cue carries lexical information.

This difference in pitch cue weighting has considerable implications for long-term L2 tone acquisition, an important topic in Mandarin pedagogy (see Yang, 2015 and Zhang, 2018 for recent reviews). For L1 English-L2 Mandarin learners who heavily weight pitch height, tones with similar F0 onsets are often confused for one another. Figure 1 shows how Tone 1 and Tone 4 both share a higher F0 onset whereas Tone 2 and Tone 3 both share a lower F0 onset. As a result, beginner learners often struggle to discriminate Tone 1 from Tone 4, and Tone 2 from Tone 3 (e.g. Hao, 2012, 2018).

Accurate L2 perception is further complicated by contextual and speaker variability, both of which can affect F0 onset (Moore & Jongman, 1997; Xu, 1997). Even after extended L2 classroom experience, many adult learners continue to perseverate on pitch height and therefore struggle to accurately discriminate and categorize tones in a native-like manner (Hao, 2012, 2018; Wang, Spence, Jongman, & Sereno, 1999; Wiener, 2017; Wiener & Lee, 2020). Yet, there is one group of non-native listeners who generally demonstrate superior tone perception: musicians.

2 Music-language interaction

For many musicians, the ability to perceive musical pitch appears to affect the perception (and production) of linguistic pitch (Bradley, 2013, 2016; Gottfried, 2007; Li & DeKeyser, 2017; Marie et al., 2011; Pfordresher & Brown, 2009; Schön, Magne, & Besson, 2004). Recent neuroimaging studies even reveal that music experience affects specific subcortical and cortical structures in the brain (Bidelman, Hutka, & Moreno, 2013; Chandrasekaran, Krishnan, & Gandour, 2009; Krishnan, Xu, Gandour, & Cariani, 2005; Wong et al., 2007). These findings suggest that experience with musical pitch helps (non-tonal L1) musicians weight linguistic pitch movement cues more heavily than non-musicians. This claim is in line with the theoretical proposal OPERA, which states that music and language interact via shared neural architecture serving similar perceptual (acoustic, categorical, and hierarchical) features (Patel, 2008, 2011, 2014). The acronym OPERA refers to the Overlap of neural architecture (for the current investigation, especially the auditory brainstem), combined with four specific task conditions known to modulate learning: Precision of categories (musical pitch categories are based on smaller degrees of precision than are typical for linguistic pitch); Emotional engagement with stimuli (facilitated by the emotional content of music); Repetition of exposure and practice of pitch categories over time; Attention during the learning process involving focused, explicit musical practice (as opposed to passive exposure and acquisition). Thus, we expect that the addition of musical interval training may benefit the learning of lexical tones because musical training involves the learning of pitch relationships similar to those involved in tone. Importantly, music learning is somewhat more attentive and repetitive because it focuses only on pitch, while language learning simultaneously involves lexical, syntactic, and pragmatic factors. Moreover, musical pitch categories are more precise, in that they are distinguished by pitch differences as small as one semitone, while tones are usually distinguished by 2–3 semitones, as well as by contour (Maddieson, 1991).

The use of music to aid child L1 linguistic abilities and adult L2 language acquisition is well documented (e.g. Abbott, 2002; François et al., 2013; Jolly, 1975; Kraus & Chandrasekaran, 2010; Richards, 1969; Tallal & Gaab, 2006). Considerable research has argued that introducing music to a language learning context provides not only linguistic benefits but also motivational and affective benefits, as learning through music can help reduce learners' anxiety and increase their self-confidence (e.g. Legg, 2009; Richards &

Rodgers, 2014). Moreover, the Emotional component of the OPERA hypothesis assumes that emotional engagement increases learning, so the inclusion of music in the curriculum can further increase emotional connections to linguistic content learning (Patel, 2011). With respect to L2 listening, previous research has explored how music can be incorporated into classroom learning to bolster perception of non-native segments and prosodic structures along with higher-level lexical and syntactic patterns (e.g. Engh, 2013; Jolly, 1975; Richards, 1969). This has extended into Mandarin tone research, which has meticulously investigated the effects of prior music and linguistic experience on tone perception (e.g. Bowles, Chang, & Karuzis, 2016; Chang, Hedberg, & Wang, 2016; Cui & Kuang, 2019; Lee & Hung, 2008; Lee & Lee, 2010; Marie et al., 2011; Zhao & Kuhl, 2015).

Yet, to our knowledge, no study has directly compared the effects of short-term musical pitch training vis-à-vis the effects of short-term linguistic pitch training. It may be that the previously reported perceptual advantage of musicians can quickly be attained through short-term musical learning. If so, musical training could lead to even greater gains when combined with structured L2 classroom learning.

II The present study

We extend previous research into the effectiveness of music on language learning by exploring how musical pitch training affects the perception of linguistic pitch. We establish an L1 Mandarin perception baseline and test three groups of L1 English adults using a pre-, post-test design with an 8-week pitch training intervention between the two tests. We define these three non-native groups by their pitch training routine: music-only (+Music), classroom-only (+L2), or simultaneous classroom and music (+L2+Music). These groups, therefore, simulate the typical adult learner of a musical instrument (+Music), of a foreign language within a classroom setting (+L2), and of both an instrument and language (+L2+Music).

We measure each group's sensitivity to Mandarin tone and weighting of pitch height and pitch movement cues. To evaluate how cue weighting changes as a function of pitch training, we make use of multidimensional scaling (Carroll & Arabie, 1980). Multidimensional scaling allows us to visualize and calculate how pitch cues are weighted across our different groups. Specifically, we use Individual Differences Scaling (INDSCAL), which assumes that the perceptual distance between stimuli reflects the time taken to discriminate between sounds (Carroll & Chang, 1970; Gandour, 1978; Gandour, & Harshman, 1978). With respect to Mandarin, INDSCAL represents the four tones in Euclidian space and allows for a direct interpretation without reconfiguration (see Chandrasekaran et al., 2010; Chandrasekaran, Gandour, & Krishnan, 2007). Cue weighting patterns can also be analysed statistically for group means (and their change from pre- to post-test) using two dimensions: pitch movement and pitch height.

III Methods

1 Participants

Thirty L1 English speakers participated in the experiment. All participants were young adults with normal hearing enrolled at one of two American universities (mean age = 20.9;

SD = 2.7). The Language Experience and Proficiency Questionnaire (LEAP-Q: Marian, Blumenfeld, & Kaushanskaya, 2007) and supplemental music-related questions were used to assess language background and music training. No participant self-reported proficiency in an L2 and all participants had fewer than five years music instrument training with no current music training. Whereas all such music training cutoffs are arbitrary, five years provided a flexible standard for the inclusion of participants with some musical experience in childhood but who had not achieved high levels of musical skill (as is common among many children in the U.S.), along with those who had never been explicitly instructed in aural skills.

Participants were assigned to one of three groups (n = 10/group) defined by their training procedure: +L2, +Music, and +L2+Music. The participants in the +L2 and +L2+Music groups were drawn from two sections of Elementary Mandarin with the same Mandarin instructor and curriculum (participants in each section took part in each condition). The participants in the +Music group were not enrolled in a foreign language class and had no previous experience with a tonal language. An additional 10 L1 Mandarin speakers from northern mainland China participated as the L1 baseline group (mean age = 21.3; SD = 2.3). These participants were selected because they self-reported being exposed only to the Mandarin variety. All participants were young adults drawn from the same university populations as the L1 English participants and thus spoke English as an L2. Participants in this group neither reported speaking an additional language nor having more than five years music training. Mean years of music training did not differ across the four groups $F(3, 36) = 0.57, p = .64$. All L1 English participants were either paid or given class credit for their time; all L1 Mandarin participants volunteered their time to help with the study.

2 Procedure

All L1 English participants took part in a pre- and post-test roughly 8-10 weeks apart. The L1 Mandarin participants only performed the test once as a baseline. The test involved two listening tasks presented over headphones in a quiet lab space. Participants first took part in the Tonometric Adaptive Pitch Test (Mandell, 2018), which identified each listener's pitch perception threshold using synthesized pure tones. Participants identified whether the second of two pure tones was higher or lower in pitch via button press. After completing the Tonometric task, participants took part in a speeded AX-discrimination task. Stimuli consisted of the Mandarin syllable *yu* spoken with the four tones by one male and one female native speaker: *yu1, yu2, yu3, yu4*. All stimuli were recorded in a sound-attenuated chamber, sampled at 44.1 kHz, and normalized to durations of 400 ms (female) and 450 ms (male). This syllable *yu* was chosen because the vowel allowed for straightforward F0 manipulation (see Xu, 1997). Stimuli creation followed the established approach outlined in Chandrasekaran et al. (2010) and Xu (1997) using Praat's (Boersma & Weenink, 2019) pitch synchronous overlap and add method to construct the F0 contours of each tone.

In the AX task, participants heard two *yu* utterances spoken by the same speaker and had to decide whether they had heard the same or different sounds. Participants were told to respond via button press as quickly and accurately as possible. A 1-second time-out period and 500 millisecond interstimulus interval was used. Participants heard 192 trials across two blocks using E-Prime (Version 2.0; Psychology Software Tools, 2012). Trials were randomized and had an equal number of male/female trials and same/different trials.

The pre-test took place during the first week of the L2 Mandarin class. Between the two testing sessions, the +L2 and +L2+Music participants took part in 3.5 classroom hours per week, along with roughly 7-10 self-reported hours of weekly outside preparation. Participants were explicitly taught the four tonal contours and their lexical role, repeatedly exposed to visualizations of the four tone contours through *pinyin* tone markings along with the tone number, and assigned regular classroom speaking and listening exercises. The +Music and +L2+Music participants took part in a commercially available computer-based musical training program, EarMaster. The music training involved exercises that helped listeners identify the structural elements of music (e.g. intervals and chords). For example, participants were trained to recognize pairs of consecutive notes differing in pitch (melodic intervals) in interval discrimination (e.g. 'which interval is larger?') and interval naming tasks (e.g. 'is this a major second or a major third?'). Participants were trained twice a week (30 minutes per session) for 8 consecutive weeks, with each week's training becoming progressively more difficult through the inclusion of a greater number of more similar intervals (e.g. week 1: major second vs. octave; week 8: perfect fourth vs. perfect fifth).

3 Data analysis

Individual Tonometric scores were used to calculate group means at pre- and post-tests. D-prime (d′) was calculated as a measure of sensitivity and to account for AX-discrimination response bias (Macmillan & Creelman, 2004). Individual d′ values were used to calculate group means at pre- and post-tests. Because the L1 Mandarin group only performed the two listening tasks once, testing session was not used as a between participant variable and thus a separate one-way between participants parametric ANOVA with equal variance (with Tukey HSD post-hoc comparisons) was used to examine means at each test. Parametric paired t-tests (with Bonferroni correction) were used to confirm whether L1 English group means changed from pre- to post-test. Normality of the data (including Tonometric scores) was confirmed using the Shapiro–Wilk test, histograms, and quantile–quantile plots.

Multidimensional scaling was carried out by creating individual tone matrices for each participant at pre- and post-tests for both the female and male stimuli. Distance estimates were calculated by taking the normalized inverse of response time for each of the six tone pair comparisons (e.g. Tone 1-Tone 2, Tone 1-Tone 3, etc.). INDSCAL analyses of these dissimilarity matrices were performed at two dimensions corresponding to pitch height and pitch movement (see Chandrasekaran et al., 2010 for additional details). Individual weights (averaged across female/male stimuli) were used to calculate group means at pre- and post-tests and analysed using a two-way parametric ANOVA with group, test, and their interaction as independent variables.

Table 1. Tonometric and sensitivity (d′) means and standard deviations by group and test.

Group	Tonometric (SD)		d′ (SD)	
L1 Mandarin	5.31 (2.72)		4.57 (.35)	
	Test:			
	Pre	Post	Pre	Post
+L2	8.98 (7.46)	7.75 (7.37)	3.54 (.51)	3.77 (.94)
+L2+Music	9.66 (7.57)	8.51 (7.05)	3.51 (1.00)	4.26 (.60)
+Music	12.78 (7.80)	9.30 (6.58)	3.58 (.79)	3.72 (.74)

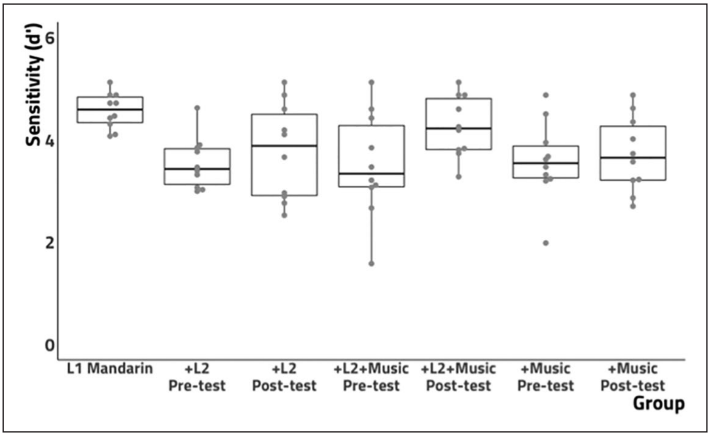

Figure 2. Mean d-prime results by participant and group.

IV Results

Table 1 presents Tonometric and d′ group means with standard deviations. Pure pitch perception as measured by Tonometric did not differ across the groups at pre-test, $F(3,36) = 2.07$, $p = .12$, or post-test, $F(3, 36) = 0.96$, $p = .42$. No L1 English group showed a significant change in pure pitch perception from pre- to post-test (*p*s $> .1$). Thus, all four groups demonstrated roughly similar mean pitch thresholds and no training modality significantly altered this mean threshold for the three L1 English groups. As a result, any observed change in AX-discrimination performance must reflect how pitch cues were weighted.

Figure 2 shows individual d′ data points over group box plots. A group difference in mean d′ was found at the pre-test, $F(3, 36) = 5.09$, $p = .005$, $\eta_p^2 = .30$. Post-hoc Tukey comparisons revealed that the L1 Mandarin group had a significantly higher mean d′ than all three L1 English groups (*p*s $< .05$) while the three L1 English groups did not differ in pre-test mean d′ (*p*s $> .9$).

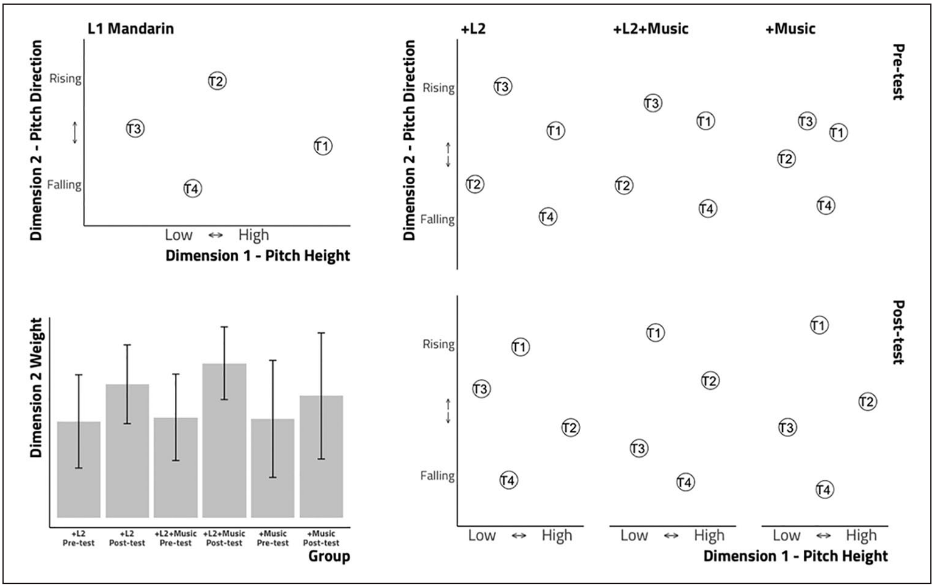

Figure 3. Mean INDSCAL configuration for L1 Mandarin (top left) and L1 English listeners (right). Mean change in Dimension 2 weights by group (bottom left).

At the post-test, a group difference in mean d′ was once again found, $F(3, 36) = 3.46$, $p = .03$, $\eta_p^2 = .22$. Post-hoc Tukey comparisons revealed that the L1 Mandarin group had a higher mean d′ than the +Music group ($p = .04$) and a marginally higher mean d′ the +L2 group ($p = .07$). The L1 Mandarin and +L2+Music group did not differ in mean d′ at post-test ($p = .74$). All three L1 English groups did not differ in mean d′ at post-test (ps $> .30$). Neither the +Music group ($t(9) = 0.56, p = .59$) nor the +L2 group ($t(9) = .66, p = .52$) differed in their mean d′ across pre- and post-tests. In contrast, the +L2+Music group had a higher mean d′ at post-test compared to their pre-test ($t(9) = 2.85, p = .02, d = .9$).

Figure 3 plots each L1 English group's INDSCAL configuration at pre- and post-tests (right) along with the L1 Mandarin group's configuration (top left). The L1 Mandarin group's space resembled the configuration reported in Chandrasekaran et al. (2010) and appropriately captured the pitch movement of the four tones. The three L1 English groups shared roughly similar configurations at pre-test in which Tone 1 and 4 separated from Tone 2 and 3 in terms of starting pitch height (Dimension 1). The L1 English groups' configuration at pre-test did not align with pitch movement (Dimension 2).

After training, all three L1 English groups demonstrated roughly similar configurations at the post-test. The L1 English participants' configuration differed from that of L1 Mandarin listeners primarily for Tone 1 and 2. The L1 English listeners appeared to treat Tone 1 as having the highest pitch movement (rather than Tone 2, as the L1 Mandarin listeners did). For both the +Music and +L2+Music group, the Dimension

2 configuration reflects the typical F0 endpoint for the four tones with Tone 1 having the highest endpoint and Tone 4 having the lowest endpoint.

The pre- and post-test Dimension 2 weights are plotted in Figure 3 (bottom left; error bars represent 95% confidence intervals) and show that all three groups increased their mean weighting of this dimension. A main effect of Test was found, $F(1, 54) = 4.78$, $p = .03$, $\eta_p^2 = .08$, indicating that overall participants weighted Dimension 2 greater at post-test compared to pre-test. However, a null effect of Group, $F(2, 54) = 0.34$, $p = .71$, and a null two-way interaction were found, $F(2, 54) = 0.26$, $p = .77$, indicating that this result was an aggregate effect across all three groups.

V Discussion

This study examined how short-term learning of musical pitch, linguistic pitch, and their combination affect cue weighting of pitch movement and pitch height. This design allowed for an examination of whether musical training effects can be exploited to improve linguistic abilities in the general population, particularly those engaged in L2 classroom learning of a tonal language. We report four findings from our study.

First, we found that L1 Mandarin speakers were more sensitive to Mandarin tones and weighted pitch movement greater than L1 English speakers at the pre-test. This finding is in line with previous research on how the role of F0 in a listener's L1 affects how F0 cues are weighted in a non-native language, particularly when perceiving a non-native tonal language (e.g. Chandrasekaran et al., 2010; Francis, Ciocca, Ma, & Fenn, 2008; Peng et al., 2010; Wayland & Guion, 2004). Moreover, this finding confirms the perceptual challenges non-tonal L1 listeners face during L2 tone acquisition (e.g. Hao, 2012, 2018; Pelzl, 2019).

Second, we found that an 8-week musical training program caused L1 English listeners to shift their weight of pitch movement cues and reconfigure their perceptual space (Figure 3). Yet, despite these gains, participants trained only on musical pitch were still less sensitive to Mandarin tone than L1 Mandarin listeners at the post-test and did not show a significant increase in tone sensitivity across tests. This suggests that short-term musical training can benefit listeners' perception of linguistic pitch, however, the previously reported perceptual advantages of musicians in tone perception and learning, and presumably any measurable changes in neuroplasticity (e.g. Chandrasekaran et al., 2007, 2009; Nan et al., 2018; Wong et al., 2007), require extended musical training beyond a short-term two month training routine. Thus, the amount of musical training undertaken by our learners – or the level of proficiency they achieved – was insufficient to produce robust changes in pitch perception akin to those observed in previously reported studies on trained musicians (e.g. Lee & Hung, 2008; Lee & Lee, 2010; Zhao & Kuhl, 2015). We also note that a participant's age may play a factor as our participants were all college aged adults. Previous short-term neuro-music studies typically examined children (e.g. Fujioka et al., 2006; Moreno et al., 2009). Whether children would show different changes in pitch perception using our approach remains an open question.

Third, we found that structured L2 Mandarin classroom training resulted in a reconfiguration of +L2 participants' perceptual space and an increase in pitch movement weights. Like the +Music group, the +L2 group did not significantly improve their tone sensitivity across tests, though we note at the post-test the L1 Mandarin group was only marginally more sensitive to tone than the +L2 group. These results corroborate previous findings that show L2 classroom learning improves non-native listeners' perception of Mandarin tone, though learners often plateau and native-like performance is certainly not the norm at lower proficiency levels (e.g. Hao, 2012; Wang et al., 1999; Wiener, Lee, & Tao, 2019).

Fourth, we found that L2 classroom instruction combined with short-term musical training resulted in a reconfiguration of +L2+Music participants' perceptual space, specifically, an increase in pitch movement weights. Unique to the +L2+Music group, sensitivity to tone at post-test was equal to that of our L1 Mandarin participants. This improvement represented a significant increase in sensitivity between the pre- and post-test and serves as initial evidence for synergy between the effects of musical and linguistic input on cue weighting and perceptual performance in an L2.

Taken together, our findings strengthen the claim that music and language interact in advantageous ways for language learners. The +Music group spent a maximum of eight hours exposed to the interval stimuli, but the majority of this training was spent actively engaged in a perceptual learning task. In contrast, the +L2 group spent over 30 classroom hours exposed to Mandarin (and an estimated 70-100 hours self-studying outside the classroom), but because the classroom learning experience included comprehensive study of vocabulary, grammar, and writing (in addition to speaking and listening), much of this experience was passive exposure. When musical training was combined with regular +L2 classroom input, a robust improvement in tonal sensitivity was observed between the pre- and post-tests. This intervention serves as an approximation of OPERA's aspects of musicianship, particularly the Overlap of auditory neural mechanisms and Precise, Repetitive, and Attentional task design (Patel, 2008, 2011, 2014). Our training program involved precise musical pitch categories and required participants' attention during a repetitive learning process spaced over the course of the term. As a result, learning musical pitch categories appeared to contribute to learning linguistic pitch categories.

Pedagogically, these results suggest that short-term musical training can be beneficial for beginner L2 Mandarin learners, especially those who struggle to accurately discriminate Tone 1 from Tone 4 and Tone 2 from Tone 3. With the proliferation of online learning, music/ear training courses are now freely available to the wider public. Even language instructors without a music background may capitalize on these courses by assigning students supplementary online music/ear training similar to our approach with EarMaster. Short-term musical training may also be beneficial for intermediate and advanced learners who continue to struggle with tone perception, though additional research is needed to evaluate this claim.

We note, however, that the simple nature of the training tasks meant that the EarMaster sessions did not engage participants emotionally; most participants reported at the post-test that the training was not very engaging, which should be a consideration for future research and for any attempts to integrate music into language curricula. If the musical training routine can be more engaging, future pedagogical studies may explore whether sequential training in which learners first complete short-term music training prior to linguistic training results in even greater perceptual gains. Additional research may explore whether the type of training – specifically singing versus listening – can create a connection between pitch perception and vocal pitch control to further improve Mandarin tone production, which beginner L2 learners also struggle to acquire (e.g. Wang, Jongman, & Sereno, 2003; Wiener, Chan, & Ito, 2020). Future work will also need to better control previous musical experience and ability, especially rhythm, which appears to play an important yet often overlooked role in language acquisition (e.g. Langus, Mehler, & Nespor, 2017; Nazzi & Ramus, 2003).

VI Conclusions

The present study demonstrated that cue weighting of challenging L2 speech sounds, such as lexical tone, can take time and rarely results in native-like performance for beginner learners. Yet, additional non-linguistic, musical training may prove beneficial. By itself, short-term musical pitch training was as effective as structured L2 classroom learning in terms of sensitivity to tone and cue weighting of pitch. This finding hints at the potentially transformative power of music to affect speech while also underscoring the challenges of L2 tone acquisition.

Funding

This research received financial support from Penn State Brandywine.

ORCID iD

Seth Wiener https://orcid.org/0000-0002-7383-3682

References

Abbott, M. (2002). Using music to promote L2 learning among adult learners. *TESOL Journal*, *11*, 10–17.

Abramson, A.S., & Whalen, D.H. (2017). Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics*, *63*, 75–86.

Bidelman, G.M., Hutka, S., & 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*, e60676.

Boersma, P., & Weenink, D. (2019) *Praat: Doing phonetics by computer* [computer program]. Available at: praat.org (accessed October 2020).

Bowles, A.R., Chang, C.B., & Karuzis, V.P. (2016). Pitch ability as an aptitude for tone learning. *Language Learning*, *66*, 774–808.

Bradley, E.D. (2013). Pitch perception in lexical tone and melody. *Reviews of Research in Human Learning and Music*, *1*, 1–26.

Bradley, E.D. (2016). Phonetic dimensions of tone language effects on musical melody perception. *Psychomusicology: Music, Mind, and Brain*, *26*, 337–345.

Bradlow, A.R., Akahane-Yamada, R., Pisoni, D.B., & Tohkura, Y.I. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, *61*, 977–985.

Carroll, J.D., & Arabie, P. (1980). Multidimensional scaling. *Annual Review of Psychology*, *31*, 607–649.

Carroll, J.D., & Chang, J.J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart-Young' decomposition. *Psychometrika*, *35*, 283–319.

Chandrasekaran, B., Gandour, J.T., & Krishnan, A. (2007). Neuroplasticity in the processing of pitch dimensions: A multidimensional scaling analysis of the mismatch negativity. *Restorative Neurology and Neuroscience*, *25*, 195–210.

Chandrasekaran, B., Krishnan, A., & Gandour, J.T. (2009). Relative influence of musical and linguistic experience on early cortical processing of pitch contours. *Brain and Language*, *108*, 1–9.

Chandrasekaran, B., Sampath, P.D., & Wong, P.C. (2010). Individual variability in cue-weighting and lexical tone learning. *The Journal of the Acoustical Society of America*, *128*, 456–465.

Chang, D., Hedberg, N., & Wang, Y. (2016). Effects of musical and linguistic experience on categorization of lexical and melodic tones. *The Journal of the Acoustical Society of America*, *139*, 2432–2447.

Cui, A., & Kuang, J. (2019). The effects of musicality and language background on cue integration in pitch perception. *The Journal of the Acoustical Society of America*, *146*, 4086–4096.

Engh, D. (2013). Why use music in English language learning? A survey of the literature. *English Language Teaching*, *6*, 113–127.

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, *26*, 551–585.

Escudero, P., Benders, T., & Lipski, S.C. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal of Phonetics*, *37*, 452–465.

Espy-Wilson, C.Y. (1992). Acoustic measures for linguistic features distinguishing the semivowels /wjrl/ in American English. *The Journal of the Acoustical Society of America*, *92*, 736–757.

Flege, J.E., Yeni-Komshian, G.H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory & Language*, *41*, 78–104.

Francis, A.L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, *36*, 268–294.

François, C., Chobert, J., Besson, M., & Schön, D. (2013). Music training for the development of speech segmentation. *Cerebral Cortex*, *23*, 2038–2043.

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

Gandour, J.T. (1978). Perceived dimensions of 13 tones: A multidimensional scaling investigation. *Phonetica*, *35*, 169–179.

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

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds 'L' and 'R'. *Neuropsychologia*, *9*, 317–323.

Gottfried, T.L. (2007). Effects of musical training on learning L2 speech contrasts. In Bohn, O.-S., & M.J. Munro (Eds.) *Language experience in second language speech learning* (pp. 221–237). Amsterdam: John Benjamins.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge / New York: Cambridge University Press.

Hao, Y.C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics*, *40*, 269–279.

Hao, Y.C. (2018). Second language perception of Mandarin vowels and tones. *Language and Speech*, *61*, 135–152.

Ho, A.T. (1976). The acoustic variation of Mandarin tones. *Phonetica*, *33*, 353–367.

Holt, L.L., & Lotto, A.J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America*, *119*, 3059–3071.

Howie, J.M. (1976). *Acoustical studies of Mandarin vowels and tones*. Cambridge / New York: Cambridge University Press.

Ingvalson, E.M., Holt, L.L., & McClelland, J.L. (2012). Can native Japanese listeners learn to differentiate/r–l/on the basis of F3 onset frequency?. *Bilingualism: Language and Cognition*, *15*, 255–274.

Iverson, P., Kuhl, P.K., Akahane-Yamada, R., et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*, B47–B57.

Jolly, Y.S. (1975). The use of songs in teaching foreign languages. *The Modern Language Journal*, *59*, 11–14.

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, *11*, 599–605.

Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research*, *25*, 161–168.

Langus, A., Mehler, J., & Nespor, M. (2017). Rhythm in language acquisition. *Neuroscience & Biobehavioral Reviews*, *81*, 158–166.

Lee, C.Y., & Hung, T.H. (2008). Identification of Mandarin tones by English-speaking musicians and nonmusicians. *The Journal of the Acoustical Society of America*, *124*, 3235–3248.

Lee, C.Y., & Lee, Y.F. (2010). Perception of musical pitch and lexical tones by Mandarin-speaking musicians. *The Journal of the Acoustical Society of America*, *127*, 481–490.

Legg, R. (2009). Using music to accelerate language learning: An experimental study. *Research in Education*, *82*, 1–12.

Li, M., & DeKeyser, R. (2017). Perception practice, production practice, and musical ability in L2 Mandarin tone-word learning. *Studies in Second Language Acquisition*, *39*, 593–620.

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

MacKain, K.S., Best, C.T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, *2*, 369–390.

Macmillan, N.A., & Creelman, C.D. (2004). *Detection theory: A user's guide*. Mahwah, NJ: Lawrence Erlbaum.

Maddieson, I. (1991). Tone spacing. *York Papers in Linguistics*, *15*, 149–175.

Mandell, J. (2018). *Tonometric adaptive pitch test*. Available at: http://jakemandell.com/adaptivepitch (accessed November 2020).

Marian, V., Blumenfeld, H.K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*, 940–967.

Marie, C., Delogu, F., Lampis, G., Belardinelli, M.O., & Besson, M. (2011). Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. *Journal of Cognitive Neuroscience*, *23*, 2701–2715.

Miyawaki, K., Jenkins, J.J., Strange, W., et al. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, *18*, 331–340.

Moore, C.B., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *The Journal of the Acoustical Society of America*, *102*, 1864–1877.

Moreno, S., Marques, C., Santos, A., et al. (2009). Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. *Cerebral Cortex*, *19*, 712–723.

Nan, Y., Liu, L., Geiser, E., et al. (2018). Piano training enhances the neural processing of pitch and improves speech perception in Mandarin-speaking children. *Proceedings of the National Academy of Sciences*, *115*, E6630–E6639.

Nazzi, T., & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. *Speech Communication*, *41*, 233–243.

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

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

Patel, A.D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research*, *308*, 98–108.

Pelzl, E. (2019). What makes second language perception of Mandarin tones hard?: A non-technical review of evidence from psycholinguistic research. *Chinese as a Second Language. The Journal of the Chinese Language Teachers Association, USA*, *54*, 51–78.

Peng, G., Zheng, H.Y., Gong, T., et al. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics*, *38*, 616–624.

Perrachione, T.K., Lee, J., Ha, L.Y., & Wong, P.C. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, *130*, 461–472.

Pfordresher, P.Q., & Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attention, Perception, & Psychophysics*, *71*, 1385–1398.

Psychology Software Tools. (2012). *E-Prime: Version 2.0* [computer program]. Pittsburgh, PA: Psychology Software Tools.

Richards, J. (1969). Songs in language learning. *TESOL Quarterly*, *3*, 161–174.

Richards, J.C., & Rodgers, T.S. (2014). *Approaches and methods in language teaching*. Cambridge / New York: Cambridge University Press.

Saito, K. (2013). The acquisitional value of recasts in instructed second language speech learning: Teaching the perception and production of English /ɹ/ to adult Japanese learners. *Language Learning*, *63*, 499–529.

Saito, K. (2015). Communicative focus on second language phonetic form: Teaching Japanese learners to perceive and produce English /ɹ/ without explicit instruction. *Applied Psycholinguistics*, *36*, 377–409.

Schertz, J., & Clare, E.J. (2020). Phonetic cue weighting in perception and production. *Wiley Interdisciplinary Reviews: Cognitive Science*, *11*, e1521.

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

Tallal, P., & Gaab, N. (2006). Dynamic auditory processing, musical experience and language development. *Trends in Neurosciences*, *29*, 382–390.

Toscano, J.C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science*, *34*, 434–464.

Toscano, J.C., McMurray, B., Dennhardt, J., & Luck, S.J. (2010). Continuous perception and graded categorization: Electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological Science*, *21*, 1532–1540.

Wang, Y., Jongman, A., & Sereno, J.A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America*, *113*, 1033–1043.

Wang, Y., Spence, M.M., Jongman, A., & Sereno, J.A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, *106*, 3649–3658.

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

Wiener, S. (2017). Changes in early L2 cue-weighting of non-native speech: Evidence from learners of Mandarin Chinese. In Proceedings of *INTERSPEECH* (pp. 1765–1769).

Wiener, S., & Lee, C.Y. (2020). Multi-talker speech promotes greater knowledge-based spoken Mandarin word recognition in first and second language listeners. *Frontiers in Psychology*, *11*, 214.
Wiener, S., Chan, M.K.M., & Ito, K. (2020). Do explicit instruction and high variability phonetic training improve nonnative speakers' Mandarin tone productions? *The Modern Language Journal*, *104*, 152–168.
Wiener, S., Lee, C.Y., & Tao, L. (2019). Statistical regularities affect the perception of second language speech: Evidence from adult classroom learners of Mandarin Chinese. *Language Learning*, *69*, 527–558.
Wong, P.C., & Perrachione, T.K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*, 565–585.
Wong, P.C., Skoe, E., Russo, N.M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, *10*, 420–422.
Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, *25*, 61–83.
Yang, B. (2015). *Perception and production of Mandarin tones by native speakers and L2 learners*. Berlin: Springer Berlin Heidelberg.
Ylinen, S., Uther, M., Latvala, A., et al. (2010). Training the brain to weight speech cues differently: A study of Finnish second-language users of English. *Journal of Cognitive Neuroscience*, *22*, 1319–1332.
Zhang, H. (2018). *Second language acquisition of Mandarin Chinese tones: Beyond first-language transfer*. Leiden / Boston, MA: Brill.
Zhao, T.C., & Kuhl, P.K. (2015). Effect of musical experience on learning lexical tone categories. *The Journal of the Acoustical Society of America*, *137*, 1452–1463.