

Research Article

Effect of Linguistic and Musical Experience on Distributional Learning of Nonnative Lexical Tones

Jia Hoong Ong,^a Denis Burnham,^a Paola Escudero,^a and Catherine J. Stevens^a

Purpose: Evidence suggests that extensive experience with lexical tones or musical training provides an advantage in perceiving nonnative lexical tones. This investigation concerns whether such an advantage is evident in learning nonnative lexical tones based on the distributional structure of the input.

Method: Using an established protocol, distributional learning of lexical tones was investigated with tone language (Mandarin) listeners with no musical training (Experiment 1) and nontone language (Australian English) listeners with musical training (Experiment 2). Within each experiment, participants were trained on a bimodal (2-peak) or a unimodal (single peak) distribution along a continuum spanning a Thai lexical tone minimal pair. Discrimination

performance on the target minimal pair was assessed before and after training.

Results: Mandarin nonmusicians exhibited clear distributional learning (listeners in the bimodal, but not those in the unimodal condition, improved significantly as a function of training), whereas Australian English musicians did not (listeners in both the bimodal and unimodal conditions improved as a function of training).

Conclusions: Our findings suggest that veridical perception of lexical tones is not sufficient for distributional learning of nonnative lexical tones to occur. Rather, distributional learning appears to be modulated by domain-specific pitch experience and is constrained possibly by top-down interference.

In a highly structured environment, learners make use of statistical regularities in the environment to extract knowledge embedded in the input. This type of learning, termed *statistical learning*, appears to be used by learners when acquiring aspects of language, from higher order linguistic knowledge such as syntactic categories (e.g., Reeder, Newport, & Aslin, 2013), word-object mapping (e.g., Smith & Yu, 2008), and word segmentation (e.g., Saffran, Aslin, & Newport, 1996), to lower level knowledge such as phonetic categories (e.g., Maye, Werker, & Gerken, 2002). This article will focus on learning a particular type of speech sound, lexical tones,¹ from statistical regularities in the input.

Infants and adults both appear to learn speech sounds appropriate for their language environment from the frequency of items in their linguistic input. This specific form of statistical learning is termed *distributional learning* and was first investigated by Maye and colleagues (Maye & Gerken, 2000, 2001; Maye, Weiss, & Aslin, 2008; Maye et al., 2002; Yoshida, Pons, Maye, & Werker, 2010) with both infant and adult learners using stop consonants.

Using a synthesized continuum from a prevoiced through to a voiceless stop consonant (e.g., a [d]-[t]), half the learners were presented with a *bimodal distribution*, in which the modal tokens were toward each end of the continuum, whereas the other half heard a *unimodal distribution*, in which the modal tokens were from the center of the continuum. It was reasoned that if learners tracked the items that they heard, then those in the bimodal condition should perceive two separate speech sounds, which would facilitate discrimination of the minimal pair represented by the end points of the continuum, whereas those in the unimodal condition should perceive only one speech sound, which would not facilitate discrimination of the minimal pair. It was found that after exposure, learners in the bimodal condition, but not those in the unimodal condition, reliably discriminated the minimal pair, providing evidence for distributional learning. Given these findings, some researchers suggest that distributional learning may underpin perceptual attunement of speech sounds in infants (Werker, Yeung, & Yoshida, 2012).

^aThe MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Penrith, New South Wales, Australia
Correspondence to Jia Hoong Ong: jhong@ntu.edu.sg

Editor: Julie Liss

Associate Editor: Bharath Chandrasekaran

Received February 27, 2016

Revision received May 26, 2016

Accepted August 7, 2016

DOI: 10.1044/2016_JSLHR-S-16-0080

¹In tone languages, word meaning can be based on lexical tone, a phonological cue based mainly on pitch (the psychological correlate of fundamental frequency, F0). For example, in Central Thai, /k^ha33/ (a mid-level tone) means “to be stuck”, whereas /k^ha241/ (a rise-fall tone) means “to kill.” In this article, lexical tones are described using Chao values (Chao, 1930) where the numbers 1 to 5 describe the relative pitch height of the tone, with 1 being the lowest and 5 being the highest, and combinations of numbers describe the contour of the pitch over the duration of the tone.

Disclosure: The authors have declared that no competing interests existed at the time of publication.

and second-language acquisition in adults (Escudero, 2005, 2009; van Leussen & Escudero, 2015).

Distributional learning with adult learners is not as clear-cut as with infants, at least with distributional learning of vowels: Whereas infants show successful distributional learning of vowels (Wanrooij, Boersma, & van Zuijen, 2014a, 2014b), results for distributional learning of vowels by adults are mixed. For example, some researchers found no significant difference in discrimination performance of adults trained on a bimodal distribution of a Dutch vowel contrast /ɑ/-/a:/ and a control group, who listened to music during training rather than a unimodal distribution (Escudero, Benders, & Wanrooij, 2011; Terry, Ong, & Escudero, 2015; Wanrooij, Escudero, & Raijmakers, 2013). Others, on the other hand, found distributional learning that was maintained for at least 6 months using the same Dutch vowel contrast (Escudero & Williams, 2014). Such inconsistencies may be due to individual differences in attention to the distribution in general (Ong, Burnham, & Escudero, 2015a; Terry et al., 2015) or differential cue weighting (Wanrooij et al., 2013) or both. Concerning the first explanation, individual differences in attention, this is similar to (a) the finding that electrophysiological responses are larger when participants actively, as opposed to passively, attend to stimuli (e.g., Gomes et al., 2000; Shafer, Morr, Datta, Kurtzberg, & Schwartz, 2005; Tervaniemi, Just, Koelsch, Widmann, & Schröger, 2005), and (b) our previous finding that distributional learning is more readily observed when learners listen attentively to the distribution via a concurrent auditory task rather than when such task is not provided (Ong et al., 2015a).

The second possible explanation for the mixed results among adult learners is the use of differential cue weighting in perceiving the minimal pair. It may be that some acoustic cues are perceived more readily than others depending on the learner's (linguistic and/or musical) experience. For example, if a learner's native language uses duration as a phonemic cue, then it is likely that the learner will perceive a pair of nonnative speech sounds minimally contrasted by a durational cue more accurately than a learner whose native language does not use duration contrastively. Musical training may similarly prompt musicians to be more sensitive to acoustic dimensions that are important in music production and perception such as timing and pitch than nonmusicians. Indeed, this has been found to be the case (e.g., Pajak & Levy, 2014; Sadakata, van der Zanden, & Sekiyama, 2010, but see Chládková, Escudero, & Lipski, 2013). For example, Korean listeners, whose language uses duration on certain vowels contrastively, are able to discriminate nonnative consonant pairs distinguished by duration more accurately than Mandarin listeners, whose language does not use duration phonemically (Pajak & Levy, 2014).

In a distributional learning task, we propose that experience is likely to affect learners' sensitivity to the target acoustic cue, which would subsequently affect their ability to extract the statistics of the distribution based on the target acoustic cue. Indeed, some researchers suggest that

part of the variability in statistical learning arises from learners' variability in encoding/processing the input (Frost, Armstrong, Siegelman, & Christiansen, 2015), which is in line with previous findings suggesting that sensitivity to an acoustic cue prior to training influences learners' ability to learn from the distributional structure of the input (Goudbeek, Cutler, & Smits, 2008; Wanrooij et al., 2013). Furthermore, as proposed by the Second Language Linguistic Perception model (L2LP; Escudero, 2005, 2009; van Leussen & Escudero, 2015), speech sounds that are "similar" to a learner's native language inventory will be easier to acquire than those that are "new" (i.e., absent from their native language inventory), as the former requires only shifting of category boundaries, whereas the latter requires the learner to form new categories. Therefore, having extensive experience with a particular acoustic cue may benefit learners in a distributional learning task based on that target cue, not just because they can encode the items in the input more accurately, but also because learners only need to *shift* rather than *form* category boundaries.

In our previous study, we demonstrated that nontone language nonmusicians (Australian English [AusE] nonmusicians) are able to acquire lexical tones distributionally (Ong et al., 2015a). By manipulating the distribution encountered, learners who were exposed to relatively more typical exemplars of the target lexical tones (bimodal condition) were better able to discriminate lexical tones after training relative to those who were exposed to relatively more ambiguous exemplars of the target lexical tones (unimodal condition). Lexical tones provide a rich avenue to investigate whether cue sensitivity modulates distributional learning because two population groups—native listeners of a tone language (henceforth, tone language listeners) and native listeners of a nontone language who are musically trained (hereafter, musicians)—consistently show an advantage in perceiving lexical tones at the behavioral level (e.g., Alexander, Wong, & Bradlow, 2005; Burnham, Brooker, & Reid, 2014; Burnham, Ciocca, & Stokes, 2001; Burnham et al., 2015; Qin & Mok, 2012; Wayland & Guion, 2004; Wong & Perrachione, 2007) and the electrophysiological level (e.g., Chandrasekaran, Krishnan, & Gandour, 2007, 2009; Krishnan, Swaminathan, & Gandour, 2009; Marie, Delogu, Lampis, Belardinelli, & Besson, 2011). This heightened sensitivity to lexical tones may well arise from different sources because the two groups use pitch differently (in a lexical context for tone language listeners and in a musical context for nontone language musicians). In this article, we examine whether *pitch experts* (tone language listeners and nontone language musicians), who perceive lexical tones more readily, might show greater distributional learning relative to AusE nonmusicians. Furthermore, there is intriguing evidence to suggest that extensive musical training is related to better performance in statistical learning (e.g., Chobert, François, Velay, & Besson, 2014; François, Chobert, Besson, & Schön, 2013; François & Schön, 2013; Paraskevopoulos, Kuchenbuch, Herholz, & Pantev, 2012; Shook, Marian, Bartolotti, & Schroeder, 2013), which suggests that in a lexical tone distributional learning task,

musicians may exhibit an additive effect of (a) higher performance in perceiving lexical tones, plus (b) better extraction of general regularities from the input.

We first investigated whether these pitch experts are able to learn lexical tones solely from the distributional structure of the input by applying the procedure from a previous experiment on distributional learning of lexical tone (Ong et al., 2015a) to two population groups: those with additional experience with pitch contrasts based only on linguistic experience (Mandarin nonmusicians; Experiment 1) and those with additional experience with pitch contrasts based only on musical experience (AusE musicians; Experiment 2). Within each population group, participants will be trained on continuum spanning a lexical tone minimal pair with either a bimodal or a unimodal distribution. Discrimination performance on the target lexical tone will be compared before and after training. Although one may be wary of the use of a discrimination task as a means to investigate learning of speech sounds, we argue that the use of discrimination tasks is both common and valid in assessing distributional learning. First, phonetic category learning studies with infants often only use discrimination tasks (e.g., Maye et al., 2002) and most, if not all, distributional learning studies, even those with adult participants, have used discrimination tasks as a proxy to index distributional learning (e.g., Escudero et al., 2011; Escudero & Williams, 2014; Maye & Gerken, 2000; Wanrooij et al., 2013). Second, despite forming abstract categories, listeners maintain their sensitivity to within-category differences due to veridical encoding of the sound (Galle & McMurray, 2014); therefore, the traditional *categorical perception* (i.e., an inability to discriminate within-category differences and a steep identification slope across category boundaries due to warping from an abstract category) may not necessarily be observed, even for stop consonants that are traditionally perceived in a categorical manner (McMurray & Aslin, 2005). For instance, vowels are considered as separate categories even though they are typically perceived in a continuous manner—that is, within-category differences among vowels are discernible, and a shallow identification slope is commonly observed along a continuum of vowels (e.g., Fry, Abramson, Eimas, & Liberman, 1962). Thus, given the nature of lexical tones and their similarity to vowels in terms of their articulatory realization, one might expect that lexical tones, too, may be perceived continuously despite the forming of abstract lexical tone categories. In that sense, if distributional learning of lexical tones *does* occur, then discrimination performance in the unimodal condition should be based on a within-category difference (due to perceiving the continuum as a single sound), whereas discrimination performance by the bimodal condition should be based on a between-category difference (due to perceiving the continuum as two separate sounds), which should result in a higher performance than discriminating items that are of a within-category difference.

In this study, we hypothesize that both populations will show distributional learning as indexed by an improvement in discriminating lexical tones following training on

a bimodal but not on a unimodal distribution.² Given that absolute pitch (AP) and/or pitch memory may be a confound in perceiving and learning lexical tones (Burnham et al., 2014) and that tone-language listeners and musicians tend to have better pitch memory and/or AP ability than nontone language nonmusicians (Bidelman, Hutka, & Moreno, 2013; Deutsch, Henthorn, & Dolson, 2004; Deutsch, Henthorn, Marvin, & Xu, 2006), pitch memory and AP will be measured using a familiar song task and a note-naming task, respectively, and considered when analyzing participants' distributional learning performance. The degree of distributional learning (if any) exhibited by both populations will be compared qualitatively in the General Discussion with a control group (nontone language nonmusicians) from our previous study (Ong et al., 2015a) to investigate the possible facilitative effect of extensive pitch experience on distributional learning of lexical tones—that is, by comparing (a) Mandarin nonmusicians and AusE nonmusicians, and (b) AusE musicians and AusE nonmusicians, we can examine the influence of tone language experience (in the case of the former) and the influence of musical experience (in the case of the latter) to determine whether domain-general versus domain-specific pitch processing modulates distributional learning of lexical tones.

Method

Experiment 1

Participants

Participants were 50 native Mandarin listeners (17 men, 33 women; age range = 18–42 years; $M = 25.04$ years, $SD = 4.62$ years) who started learning English at the mean age of 10.33 years ($SD = 2.43$ years) in Mainland China. Some Mandarin participants reported having musical training; however, none had more than 2 years of musical experience (≤ 0.5 year, $n = 3$; 1 year, $n = 3$; 2 years, $n = 1$). All reported normal hearing. The Mandarin participants were students from Western Sydney University or the University of New South Wales, and they were paid \$15 for their participation. All participants gave written informed consent prior to participating in the experiment. The Western Sydney University Human Research Ethics Committee approved the study protocol.

Stimuli

Distributional learning task. The stimuli were identical to those in Ong et al. (2015a). Four native Thai speakers

²It should also be noted that given the current experimental design (pretest–training–posttest), some improvement is to be expected due to practice, irrespective of the distribution encountered during training (Ong et al., 2015a). However, if the unimodal condition fails to show significant improvement *despite* the pretest–training–posttest design, then this can be taken as evidence for a weakened sensitivity to lexical tones, most likely due to unimodal participants perceiving the sounds along the continuum to be from a single lexical tone and therefore performing the discrimination task by relying on within-category differences after training.

(two women) produced multiple tokens of four real Thai words:³ /k^ha33/, /k^ha241/, /na33/, and /na241/. From these, two minimal tone pairs produced by the same speaker were formed between words with the same phones but different tones (i.e., /k^ha33/-/k^ha241/ and /na33/-/na241/). The choice of this midlevel versus falling contour tone minimal pair was based on a previous study demonstrating it is the most difficult Thai tone minimal pair for nontone language listeners and tone language listeners to discriminate (Burnham et al., 2015). As there were both male and female speakers, four minimal pairs were used in this study: female /k^ha33/-/k^ha241/; female /na33/-/na241/; male /k^ha33/-/k^ha241/; and male /na33/-/na241/.

To form multiple exemplars of each minimal pair, we first chose a base sound file comparable in duration and matching for the speaker who produced that base sound file for each minimal pair. Then, we extracted the pitch contour from each member of the minimal pair chosen from other recording tokens of the same word matching for duration from the same speaker. The extracted pitch contour was then imposed on the base sound file to ensure that only the pitch contour differed between the members of each minimal pair (Ong et al., 2015a). For each minimal pair, three exemplars for each tone were created using the method described above: one of which was used as a reference exemplar and two of which were used as target exemplars in an ABX discrimination task (described in the Procedure subsection below). All four minimal pairs were subjected to the same treatment, resulting in 24 tokens of test stimuli (4 minimal pairs × 2 tones × 3 exemplars). Although the duration of each minimal pair ranged from 493 to 832 ms, the duration of the two sounds within each minimal pair was equated. The test stimuli formed a 2 × 2 factorial complex: Test syllable (/k^ha/ vs. /na/) × Test gender (female vs. male speaker).

An eight-step continuum for the male /na33/-/na241/ (using the reference exemplar of each) was formed as training stimuli, with Tone 33 as Token 1 of the continuum and Tone 241 as Token 8 (see Figure 1). This minimal pair was chosen to form the training continuum, as a previous study has demonstrated that it is the most difficult to discriminate of the four minimal pairs (Ong et al., 2015a). The intermediate tokens of the continuum were created by interpolating the pitch contour of the two end tokens using Praat (Boersma & Weenink, 2013).

To familiarize the participants with the discrimination task, a 440-Hz sinewave tone and a 440-Hz sawtooth wave tone, both 800 ms in duration, were synthesized using Praat (Boersma & Weenink, 2013) and used as practice stimuli. In addition, the sinewave tone was used as beeps during training as part of the concurrent auditory vigilance task (see below; Ong et al., 2015a).

³In Thai, /k^ha33/ and /k^ha241/ mean “to be stuck” and “to kill,” respectively, whereas /na33/ and /na241/ mean “paddy field” and “face,” respectively.

Familiar song task. The familiar song task was based on that used in previous studies (Ong, Burnham, & Escudero, 2015b; Schellenberg & Trehub, 2003). Based on a pilot study, we chose 40 popular English songs and excised and duplicated the first 5 s (i.e., the instrumental portion) of each song. Half the duplications were randomly assigned to have their pitch raised, whereas the other half had their pitch lowered. The degree of transposition of the duplications was either one or two semitones, resulting in four different sets (+1, +2, -1, and -2) with each set comprising 10 songs. To remove any artefacts due to digital manipulation, the pitch of the original excerpts was also transposed upward and then downward to the same degree.

Equipment

All tasks were presented using MATLAB 2012b (The MathWorks, Inc., Natick, Massachusetts, United States) on an Acer TravelMate P653 laptop (Acer Inc., New Taipei, Taiwan). The auditory stimuli were presented via Sennheiser HD650 headphones (Sennheiser electronic GmbH & Co. KG, Hanover, Lower Saxony, Germany) connected to an Edirol USB Audio Capture UA-25EX audio interface (Roland Corporation, Los Angeles, California, United States).

Procedure

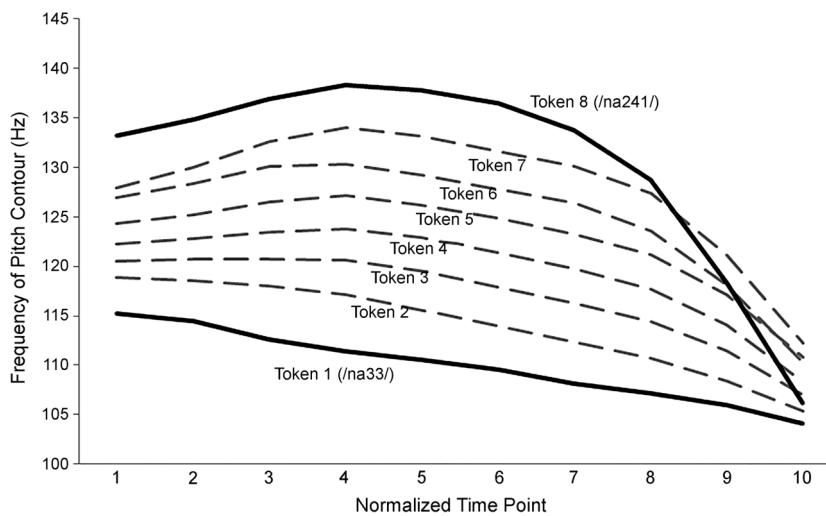
Participants completed the distributional learning task, language and musical background questionnaire, and the familiar song task in a randomized order for each participant. The general procedure for each task is described below.

Distributional learning task. Participants were randomly assigned to one of two distribution conditions: bimodal or unimodal. To familiarize participants with the format of the ABX discrimination task, four practice trials with feedback were presented. In these practice trials, sine and sawtooth waves were used. Participants indicated whether the third sound was similar to the first or the second by pressing the left shift key (A and X are similar) or the right shift key (B and X are similar). Participants were informed that they were required to respond within 1 s.

Following practice, the distributional learning task was presented with three phases: pretest, training, and posttest. During pretest and posttest, participants were asked to discriminate all four test minimal pairs in an ABX discrimination task. The first and second sounds (i.e., A and B) were always reference exemplars,⁴ whereas X was always

⁴The choice of using the end points as reference exemplars is motivated by two reasons: (a) Those stimuli are naturalistic tokens of the lexical tones, which sets our experiment apart from a previous distributional learning experiment with lexical tones that used artificial stimuli and did not find an effect (Perfors & Ong, 2012); and (b) to ensure our study is comparable with other distributional learning experiments (e.g., Escudero et al., 2011; Escudero & Williams, 2014; Ong et al., 2015a)

Figure 1. Pitch contour of the training continuum from /na33/ (Token 1) to /na241/ (Token 8). Note that the pitch contours shown here represent the tone space of the vowel, in which the first 15% and the last 15% of the vowel were excluded to remove possible effects of coarticulation from the preceding consonant and creakiness, respectively.



one of the target exemplars of the same minimal pair as the first two sounds. In both pretest and posttest, each test minimal pair was presented eight times, resulting in a total of 32 trials in each test, the order of which were randomized and not blocked. Participants were required to respond within 1 s with no replacement trials for timed-out trials.

During training, participants were instructed to listen to a sequence of sounds, and in order to ensure attentive listening throughout the training phase, participants were required to perform an auditory vigilance task similar to the one included in our previous study (Ong et al., 2015a), which required them to note on a paper response sheet whenever they heard a beep (i.e., the sine wave tone used in practice trials). A total of 32 beeps were interspersed randomly within 256 male /na/ training tokens. Although both distribution conditions were presented with all eight training tokens and the same total number of training tokens, the distribution of the training continuum differed between the two distribution conditions; bimodal condition participants heard Tokens 2 and 7 most frequently, whereas unimodal condition participants heard Tokens 4 and 5 most frequently (see Figure 2). It is crucial to note that the number of times both conditions heard Tokens 1 and 8 (i.e., the reference exemplars for the male /na33-/na241/ test minimal pair) was equal.

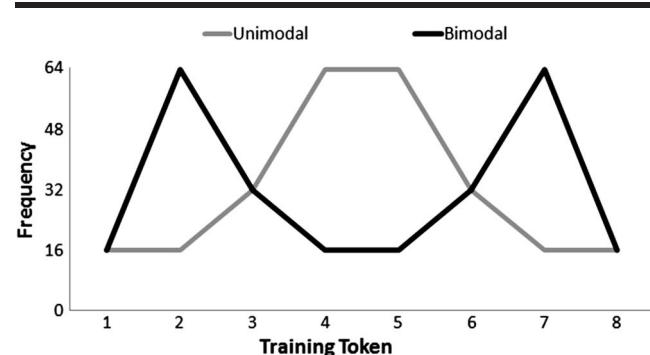
Language and musical background questionnaire. In the questionnaire, participants were asked to list all the languages they know and rate on a 5-point scale how well they (a) speak, (b) understand, (c) write, and (d) read in each of those languages, and to provide information regarding their musical background—whether they had any musical training (either self-taught, private, or formal), the age of training onset, and the duration of training in years.

Familiar song task. In the familiar song task, participants were first shown a song title and the artist who had performed the song and were asked to indicate whether they were familiar with that song. If they were, they then heard two excerpts of the song: an original (or essentially, the studio version) and a pitch-transposed version, with the order of presentation counterbalanced, and they were required to indicate which of the pair the original was. If they were unfamiliar with that particular song, they moved on to the next trial with no replacement trials. Participants were required to complete two blocks of 20 popular English songs, one with a transposition of ± 1 semitone, the other with ± 2 semitones, with presentation order within and between blocks randomized.

Results and Discussion

In the familiar song task, which measures pitch memory, the Mandarin participants were familiar with an average of 22.16 of the 40 songs ($SD = 7.78$). In percentages

Figure 2. Frequency of occurrence for each training token encountered by listeners in the unimodal and bimodal conditions.

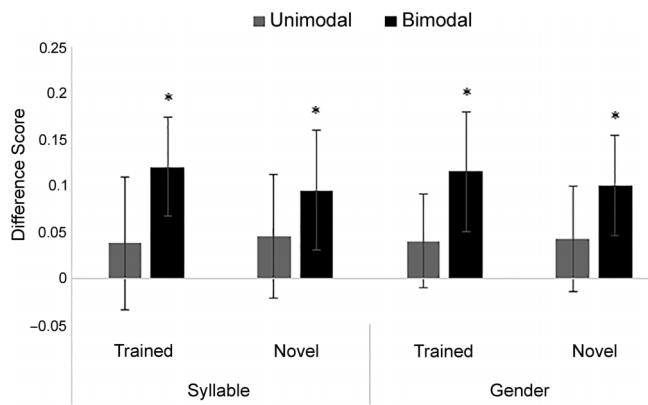


their performance on known song trials ranged from 20% to 81.48% ($M = 53.55\%$, $SD = 12.64\%$), so none achieved the AP criterion of 85% accuracy (Deutsch et al., 2006). An independent-sample t test revealed that the unimodal condition had significantly higher scores on the familiar song task than the bimodal condition, $t(48) = 2.594$, $p = .013$. We return to implications of this later.

Prior to analyzing the main results, an independent t test revealed that the two distribution conditions did not differ on pretest accuracy (unimodal: $M = .835$, $SE = .025$ vs. bimodal: $M = .781$, $SE = .025$; $t[48] = 1.508$, $p = .138$), showing that any differences between listeners in the two distribution conditions at posttest would result from the training itself. A mixed analysis of variance (ANOVA), with distribution condition (unimodal vs. bimodal) as a between-subjects factor and session (pretest vs. posttest), test syllable (trained vs. novel syllable), and test gender (trained vs. novel gender) as within-subjects factors, revealed that there were main effects of session, $F(1, 48) = 21.123$, $p < .001$, $\eta_p^2 = .306$, and test syllable, $F(1, 48) = 17.172$, $p < .001$, $\eta_p^2 = .263$: Performance was better at posttest ($M = .883$, $SE = .018$) than pretest ($M = .808$, $SE = .018$), and participants' performance on novel syllable (i.e., /k^{ha}/) test items ($M = .874$, $SE = .018$) was higher than that of trained syllable (i.e., /na/1) test items ($M = .817$, $SE = .017$). Because the trained minimal pair (male /na33/-/na241/) was chosen precisely because it is the most difficult minimal pair to discriminate, the finding that novel syllable items had generally higher accuracy than trained syllable items is therefore not surprising: It simply reflects the difficulty in discriminating the trained item. It is important to note that there was a significant Session \times Distribution Condition interaction, $F(1, 48) = 4.190$, $p = .046$, $\eta_p^2 = .080$. Simple main effect analysis revealed that for listeners in the unimodal condition, posttest scores were not significantly different from pretest scores ($M = .876$, $SE = .026$ vs. $M = .835$, $SE = .025$), $F(1, 24) = 3.293$, $p = .082$, whereas for those in the bimodal condition, posttest scores were significantly higher than pretest scores ($M = .889$, $SE = .026$ vs. $M = .781$, $SE = .025$), $F(1, 24) = 21.769$, $p < .001$, $\eta_p^2 = .476$.

In addition, we conducted a series of one-sample t tests with Holm–Bonferroni correction on posttest–pretest difference scores for each test dimension by distribution condition to determine whether participants improved after training (i.e., significantly above zero where zero means no improvement; see Figure 3). The results revealed that listeners in the unimodal condition did not improve on any of the test dimensions (trained syllable, $t[24] = 1.078$, $p = .292$, adjusted $\alpha = .05$; novel syllable, $t[24] = 1.390$, $p = .177$, adjusted $\alpha = .05$; trained gender, $t[24] = 1.619$, $p = .119$, adjusted $\alpha = .05$; novel gender, $t[24] = 1.550$, $p = .134$, adjusted $\alpha = .05$), whereas those in the bimodal condition improved on all test dimensions (trained syllable, $t[24] = 4.615$, $p < .001$, adjusted $\alpha = .0125$; novel syllable, $t[24] = 3.018$, $p = .006$, adjusted $\alpha = .05$; trained gender, $t[24] = 3.639$, $p = .001$, adjusted $\alpha = .0167$; novel gender, $t[24] = 3.843$, $p = .001$, adjusted $\alpha = .025$), thus showing generalized distributional learning over test stimuli.

Figure 3. Difference scores (posttest minus pretest) on test dimensions by distribution condition for Mandarin nonmusicians (Experiment 1). Asterisk indicates performance is significantly different from zero (i.e., no improvement) after Holm–Bonferroni correction. Error bars represent 95% confidence intervals.



Taken together, these results indicate that Mandarin listeners are able to learn nonnative lexical tones based on the distributional structure of the input; those trained on a bimodal distribution improved after training, whereas those trained on a unimodal distribution did not, over and above (a) any practice effect due to the pretest–training–posttest experimental design and (b) higher pitch memory performance by listeners in the unimodal condition relative to those in the bimodal condition. In Experiment 2, we investigate whether nontone language listeners who are musically trained (nontone language musicians), who have been shown to also be sensitive to lexical tones relative to those without musical training (e.g., Alexander et al., 2005; Burnham et al., 2014; Wong & Perrachione, 2007), will also exhibit distributional learning of lexical tones.

Experiment 2

Participants

Participants were 50 native AusE musicians (18 men, 32 women; age range = 18–38 years; $M = 26.14$, $SD = 6.01$). Almost half the AusE musicians reported being AusE monolingual speakers ($n = 24$); the rest were multilinguals, although all were naive to tone languages. The AusE musicians had at least five continuous years of musical training ($M = 14.63$, $SD = 6.17$), and all were still practicing music. The majority of the musicians were multi-instrumentalists; nine musicians listed played a single instrument. All reported normal hearing. The AusE musicians were recruited from the Sydney region, and they were paid \$20 for their participation. All participants gave their written informed consent prior to participating in the experiment. The Western Sydney University Human Research Ethics Committee approved the study protocol.

Stimuli and Equipment

The same set of stimuli and equipment as in Experiment 1 was used. In addition, stimuli for a note-naming

task were created for this experiment. The note-naming task was modeled after that of a previous study (Lee, Lekich, & Zhang, 2014). Notes ranging from C₃–B₅ were produced (36 in total), with all notes tuned to concert A (A440). Each note was produced in three different timbres: piano, cello, and sine wave tone. The piano and cello notes were produced on LogicPro 7 (Apple Inc., Cupertino, California, United States) using Alchemy's C7 piano (Apple Inc., Cupertino, California, United States) and Kontakt's Cello Ensemble audio plugins (Native Instruments, Berlin, Germany), respectively, whereas sine wave tones were produced using MATLAB R2012b (The MathWorks, Inc., Natick, Massachusetts, United States).

Procedure

The AusE musicians completed the same tasks as did the Mandarin nonmusicians in Experiment 1 with the addition of a note-naming task. In the note-naming task, the AusE musicians were told that they would hear a note and were required to name the note by typing it on a computer keyboard within 10 s. They were given three practice trials without feedback to familiarize themselves with the task. Then, participants were given three blocks of tests, each corresponding to a particular timbre (piano, cello, or sine wave tone—the order of which was randomized). The presentation of notes within each block was pseudorandomized; succeeding notes were more than an octave apart. There were no replacement trials for slow responses. Seven AusE musicians failed to complete this task (three in unimodal condition, four in bimodal condition) due to technical error.

Results and Discussion

We first analyzed the AusE musicians' performance on (a) the familiar song task, which measures pitch memory, and (b) the note-naming task, which measures AP ability. Concerning pitch memory, the AusE musicians were familiar with a mean of 26.50 of the 40 songs ($SD = 7.93$), and their percentage accuracy ranged from 29.41% to 96.55% ($M = 67.50\%$, $SD = 14.15\%$), with four scoring above the AP criterion of at least 85% accuracy (two in the bimodal condition and two in the unimodal condition; Deutsch et al., 2006). However, AP ability, as indexed by the note-naming task, ranged from 0% to 80.56% ($M = 35.16\%$, $SD = 19.28\%$), so none met the AP criterion in the note-naming task. This seemingly contradictory result in pitch memory and AP ability may be due to the task difference: The familiar song task is a two-alternative forced-choice task, whereas the note-naming task is a free-recall task, which is arguably more difficult. It is important to note that independent-sample t tests revealed no significant difference in performance between listeners in unimodal and bimodal conditions on either task (familiar song task, $t[48] = 1.230$, $p = .225$; note-naming task, $t[41] = 0.647$, $p = .521$). We also determined that listeners in the two distributions groups did not differ significantly at pretest using an independent t test (unimodal: $M = .691$, $SE = .022$ vs. bimodal: $M = .706$, $SE = .022$; $t[48] = 0.481$, $p = .633$). Then, we

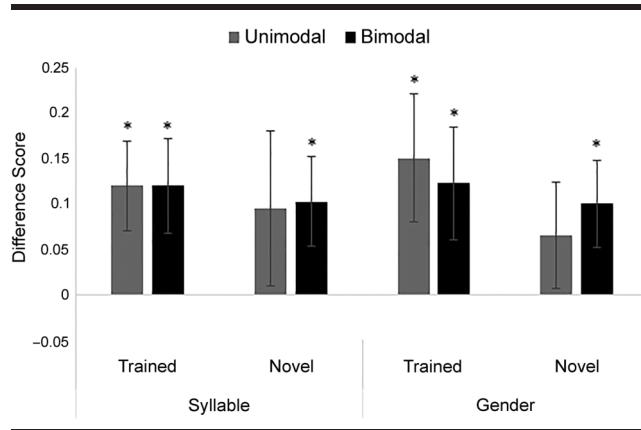
conducted the same analyses as in Experiment 1—that is, (a) a mixed ANOVA with distribution condition as a between-subjects factor and session, test gender, and test syllable as within-subjects factors; and (b) a series of one-sample t tests with Holm–Bonferroni correction on difference scores.

The ANOVA revealed main effects of session, $F(1, 48) = 51.191$, $p < .001$, $\eta_p^2 = .516$; test gender, $F(1, 48) = 162.463$, $p < .001$, $\eta_p^2 = .772$; and test syllable, $F(1, 48) = 10.756$, $p = .002$, $\eta_p^2 = .183$. In general, posttest scores ($M = .808$, $SE = .014$) were higher than pretest scores ($M = .699$, $SE = .016$); listeners scored higher on novel gender stimuli ($M = .856$, $SE = .016$) than on trained gender stimuli ($M = .651$, $SE = .014$); and participants had higher scores on trained ($M = .778$, $SE = .013$) than novel syllable ($M = .729$, $SE = .017$) test items. There was also a significant Test Gender \times Test Syllable interaction, $F(1, 48) = 19.274$, $p < .001$, $\eta_p^2 = .286$: For trained gender stimuli, participants' performance on the two syllables did not differ significantly, whereas for novel gender stimuli, participants' performance on trained syllable was higher than on novel syllable. A similar pattern of interaction was also observed in a three-way interaction between Test Gender \times Test Syllable \times Distribution Condition, $F(1, 48) = 6.939$, $p = .011$, $\eta_p^2 = .126$. For trained gender stimuli, listeners in the unimodal condition showed higher performance for novel syllable than trained syllable, whereas listeners in the bimodal condition showed similar performance across both syllables. On the other hand, for novel gender stimuli, both unimodal and bimodal condition participants showed higher performance on trained syllable than novel syllable. This seemingly counterintuitive finding of better performance on the novel items than on the trained items reflects the outcome of choosing the most difficult minimal pair of the four as the training minimal pair. More related to our hypothesis, there was no significant Session \times Distribution Condition interaction, $F(1, 48) = .015$, $p = .903$, $\eta_p^2 = .000$ —that is, the degree of difference from pretest to posttest did not differ significantly between listeners in the bimodal and unimodal conditions.

Results of one-sample t tests with Holm–Bonferroni correction (see Figure 4) revealed that listeners in the bimodal condition improved on all test dimensions (trained syllable, $t[24] = 4.755$, $p < .001$, adjusted $\alpha = .0125$; novel syllable, $t[24] = 4.240$, $p < .001$, adjusted $\alpha = .0167$; trained gender, $t[24] = 4.073$, $p < .001$, adjusted $\alpha = .025$; novel gender, $t[24] = 4.276$, $p < .001$, adjusted $\alpha = .05$), and listeners in the unimodal condition improved on two of the four dimensions (trained syllable, $t[24] = 5.018$, $p < .001$, adjusted $\alpha = .0125$; trained gender, $t[24] = 4.382$, $p < .001$, adjusted $\alpha = .0167$ vs. novel syllable, $t[24] = 2.282$, $p = .032$, adjusted $\alpha = .025$; novel gender, $t[24] = 2.279$, $p = .032$, adjusted $\alpha = .025$).

Taken together, our results indicate that AusE musicians did not show the predicted distributional learning of lexical tones; contrary to prediction, learners in the bimodal condition improved on all test dimensions and learners in the unimodal condition improved on at least two test dimensions after training. Thus, it appears that,

Figure 4. Difference scores (posttest minus pretest) on the test dimensions by distribution condition for AusE musicians (Experiment 2). Asterisk indicates performance is significantly different from zero (i.e., no improvement) after Holm–Bonferroni correction. Error bars represent 95% confidence intervals.



despite being sensitive to the target acoustic cue as shown by previous studies (e.g., Alexander et al., 2005; Burnham et al., 2014; Wong & Perrachione, 2007), sensitivity to pitch is not sufficient for distributional learning of lexical tones to occur.

General Discussion

The aim of these experiments was to determine whether the advantage in perceiving lexical tones seen among listeners with extensive experience with pitch—either in the form of lexical tones or musical training—will also be observed in the acquisition of nonnative lexical tones based on the distributional structure of the input. This was realized by administering a distributional learning task (Ong et al., 2015a) to tone language (Mandarin) nonmusicians (Experiment 1) and nontone language (AusE) musicians (Experiment 2). Mandarin nonmusicians exhibited clear distributional learning—learners in the bimodal condition showed significant improvement from pretest to posttest, whereas those in the unimodal condition did not. However, AusE musicians did not show the predicted distributional learning; learners in the bimodal condition showed significant improvement from pretest to posttest, but contrary to prediction, so did those in the unimodal condition on two of four test dimensions. Thus, when the results across both experiments are compared separately to the results of nontone language nonmusicians in our previous study (Ong et al., 2015a), it appears that experience with domain-specific pitch (i.e., experience with a tone language but not experience with musical training) modulates distributional learning of lexical tones.

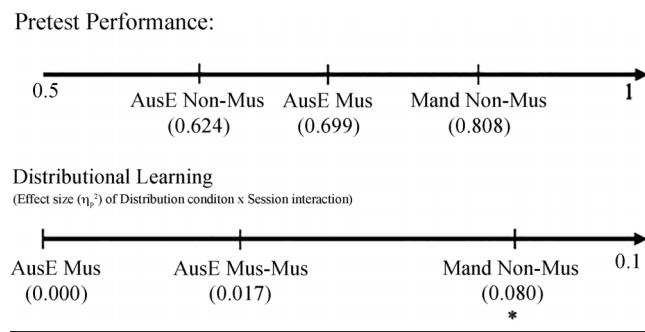
Previous studies have found that tone language listeners, in general, perceive native and nonnative lexical tones accurately (e.g., Burnham et al., 2001, 2015; Chandrasekaran et al., 2007; Qin & Mok, 2012; Wayland & Guion, 2004). Our results extend previous findings by showing that tone

language listeners can correctly learn nonnative lexical tones solely from the statistical information in the input. In fact, a qualitative comparison of the pattern of improvement/suppression exhibited by the distribution conditions with nontone language listeners (AusE nonmusicians) from our previous study (Ong et al., 2015a) revealed that the Mandarin nonmusicians here show greater distributional learning than do AusE nonmusicians; there was a larger effect size of the Session \times Distribution Condition interaction exhibited here by the Mandarin nonmusicians ($\eta_p^2 = 0.080$, a medium size effect) than by AusE nonmusicians reported in Ong et al. (2015a; $\eta_p^2 = 0.017$, a small size effect—see Figure 5). Furthermore, looking at the pattern of improvement and suppression shown in the distribution conditions, the Mandarin nonmusicians in the unimodal condition showed suppression of improvement on all four test aspects, and those in the bimodal condition showed improvement on all four test aspects. However, AusE nonmusicians in the unimodal condition in the previous study showed a suppression of improvement on three dimensions, and the bimodal condition only improved on three test dimensions.

In line with the L2LP model (e.g., Escudero, 2005; van Leussen & Escudero, 2015), we suggest that, given the same number of training tokens, Mandarin listeners benefit more⁵ than AusE nonmusicians did in our previous article (Ong et al., 2015a), presumably because the former simply shifted category boundaries to accommodate the Thai stimuli (Escudero, 2005, 2009; van Leussen & Escudero, 2015), whereas AusE nonmusicians in our previous study may have (a) formed lexical tone categories, which takes more time and is more difficult than shifting boundaries (Escudero, 2009); or (b) assimilated the Thai stimuli to their prosodic categories, which have fuzzier, less-defined boundaries (Reid et al., 2015). In either of these alternatives, the AusE nonmusicians may require more input to exhibit the same strength of distributional learning as the Mandarin participants. However, whereas second-language

⁵Another possible explanation of Mandarin listeners showing greater distributional learning than AusE nonmusicians is that all the Mandarin listeners in this study were late bilinguals compared to a mix of AusE monolinguals and bilinguals in Ong et al. (2015a). There is some indication that bilingual experience, rather than tone language experience, facilitates statistical learning of tonal words from concatenated speech (T. Wang & Saffran, 2014). However, in that study, the tonal words to be learned were cued redundantly on both the syllable and the tonal level. Thus, another interpretation of T. Wang and Saffran (2014) is that when multiple cues (syllabic and tonal) are provided to statistically learn transitional probabilities of tonal words, experience with more than one language may benefit such learning. Indeed, bilinguals outperform monolinguals in statistical learning of word-object mapping only when the object has more than one label, suggesting that bilinguals may be more flexible in tracking multiple cues of the same target item than monolinguals (Poepzel & Weiss, 2016). However, in our distributional learning task, each to-be-learned item is not cued by multiple cues; the task requires learners to simply track the speech sounds that they encounter. Therefore, it is unlikely that the observed difference between Mandarin nonmusicians and AusE nonmusicians here is due to bilingual experience.

Figure 5. Qualitative comparison of pretest performance and degree of distributional learning (indexed by the effect size of Session \times Distribution Condition interaction) in three population groups: Mandarin nonmusicians (Mand Non-Mus; Experiment 1), AusE musicians (AusE Mus; Experiment 2), and AusE nonmusicians (AusE Non-Mus; data from Ong et al., 2015a). Only the Mandarin nonmusicians showed a significant Session \times Distribution Condition interaction (as indicated by the asterisk; i.e., significant distributional learning). AusE nonmusicians showed a significant three-way interaction of Session \times Distribution Condition \times Syllable instead. AusE = Australian English.



speech perception models such as the Perceptual Assimilation Model (Best, 1995) and the Second Language Perceptual Assimilation model (Best & Tyler, 2007) would both predict that the Mandarin participants shifted their existing (native) lexical tone category boundaries to accommodate the nonnative Thai tones, the present results do not allow us to determine whether Mandarin participants do indeed shift existing native lexical tone category boundaries, or alternatively, as predicted by the L2LP (Escudero, 2005, 2009; van Leussen & Escudero, 2015), adjust the category boundaries of lexical tone that are a direct copy of their native phonological system—that is, their own native lexical tones remain unaffected. Further work is necessary to delineate the exact mechanisms that determine the present results.

Why might AusE musicians not show distributional learning (as indexed by both a lack of significant Session \times Distribution Condition interaction and the pattern of suppression of improvement by the unimodal condition), whereas Mandarin nonmusicians in Experiment 1 and AusE nonmusicians in Ong et al. (2015a) do? This is not because the musicians performed at ceiling, because the Mandarin nonmusicians had an even higher accuracy performance at pretest relative to the AusE musicians, and yet the Mandarin nonmusicians showed the predicted distributional learning (see Figure 5). The lack of distributional learning observed among the AusE musicians seems counterintuitive given that (a) the vast majority of the literature has reported that musicianship facilitates discrimination and learning of lexical tones (e.g., Alexander et al., 2005; Burnham et al., 2014; Lee et al., 2014; Wong & Perrachione, 2007), and thus, musicians should be better able to encode and process lexical tones in the input to detect the distributional structure (Frost et al., 2015); and (b) musicians are better able to extract statistical regularities from the input

than are nonmusicians in general (e.g., François et al., 2013; Paraskevopoulos et al., 2012; Schön & François, 2011; Shook et al., 2013). As indicated in Figure 5, it appears that accurate perception of lexical tones is not sufficient for distributional learning to occur.

We propose that the musicians in this study did not acquire lexical tones due to top-down interference, which has been shown to hinder distributional learning (Gulian, Escudero, & Boersma, 2007). For example, when training tokens (nonnative vowels) were explicitly labeled with orthography of the learners' native language, learners did not show distributional learning presumably due to interference from metalinguistic knowledge of the target speech sounds (Gulian et al., 2007). In a similar vein, the musicians in this study may not have treated the isolated lexical tones as speech (indeed, several musicians commented during postexperiment debriefing that the isolated Thai lexical tones sounded like sung notes), which would affect their expectations of or how they process the signal. In other words, the musicians may not need to form lexical tones to perform the task due to how they *interpret* the signal (i.e., they could just rely on their robust musical representations), and so did not show any distributional learning of lexical tones. If so, then this suggests that exposure to a simplified lexical tone distribution may not result in distributional learning if learners approach the stimuli in a musical manner.

Our conjecture is in line with previous suggestions that the functional context of the pitch signal, as interpreted by the listener, affects how the signal will be processed (Baudoin-Chial, 1986; Van Lancker & Fromkin, 1973, 1978; Y. Wang, Jongman, & Sereno, 2001). Using a dichotic listening paradigm, tone language listeners show a right ear advantage for lexical tones, but a left ear advantage for hummed version of lexical tones, whereas nontone language listeners show a left ear advantage for both types of stimuli (Van Lancker, 1980). Furthermore, nonexperienced singers show greater activation of the language network in the brain during vocal singing, whereas experienced singers recruit different brain areas from the language network during vocal singing (Wilson, Abbott, Lusher, Gentle, & Jackson, 2011), thereby suggesting that the same tasks may elicit dissociations in activation of brain areas depending on musical training. In terms of behavior, nontone language nonmusicians have been shown to have graded discrimination performance on stimuli that are music-like to speech-like—that is, best performance on violin notes with the same pitch contour as lexical tones, followed by low-pass filtered lexical tones and worst performance in naturalistic lexical tones/speech (Burnham et al., 2014). On the other hand, musicians do not show this graded response across context (Burnham et al., 2014), suggesting that nonmusicians may use different modes of processing depending on the context, whereas musicians may approach all three contexts in the same (musical) manner. In the present study and in Ong et al. (2015a), the instruction given to participants was deliberately broad: We asked them to listen to the sounds without specifying whether they were musical or linguistic. Thus, without any

contextual information, nontone language musicians may interpret the signal differently from nontone language nonmusicians, which may lead to the use of a different mode of processing. Perhaps distributional learning in musicians would be observed if a speech-mode of processing was induced during training and/or testing—for instance, training the participants using a distribution of lexical tones embedded within short phrases or mapped to objects such as in word learning tasks.

This study has only investigated the influence of extensive lexical pitch experience (Experiment 1) and extensive musical pitch experience (Experiment 2) on distributional learning of lexical tones. One might ask, how would tone-language musicians (e.g., Mandarin musicians) perform in this study? Assuming that the two forms of experience are independent and that one form of experience might overpower the other, depending on how Mandarin musicians interpret the signal, they may show similar performance to AusE musicians (i.e., they do not show distributional learning of lexical tones) or to Mandarin nonmusicians (i.e., they show distributional learning of lexical tones). However, it may be the case that both experiences provide an additive effect such that Mandarin musicians may outperform Mandarin nonmusicians. In order to fully understand the individual and joint contributions of lexical pitch and musical pitch experience, further research with tone language musicians is needed.

If distributional learning is partly modulated by experience with the relevant acoustic cue, which would affect how the learner interprets the signal—that is, whether a speech- or music-mode of processing was used—then musicians should show the largest effect of distributional learning for sung (musical) pitch categories relative to nonmusicians. The question, then, would be whether tone language experience facilitates the distributional learning of musical pitch. On the one hand, previous research suggests that tone language listeners show better performance in perceiving musical/nonspeech stimuli than nontone language listeners (e.g., Alexander, Bradlow, Ashley, & Wong, 2011; Bidelman, Gandour, & Krishnan, 2011; Krishnan et al., 2009; Pfördresher & Brown, 2009; Stevens, Keller, & Tyler, 2011). On the other hand, the results of the present study suggest that the advantage of extensive pitch experience is constrained by the manner in which the learners approach the task: If tone-language listeners approach sung stimuli as speech, then they may not show any distributional learning as they would presumably be relying on their native lexical tones to perform the task rather than acquiring the novel musical pitch based on the input. In addition, the relative ease or difficulty of acquisition may also depend on the sung stimuli themselves and how the target stimuli are perceived by the learners (i.e., whether the sung stimuli are perceived as similar or new contrasts; see Escudero, 2005, 2009; van Leussen & Escudero, 2015). Work is currently under way in our lab to address this.

In conclusion, our results indicate that correctly perceiving an acoustic cue is not sufficient for distributional learning to occur. When we compare the results of AusE

nonmusicians from our previous study to (a) Mandarin nonmusicians (Experiment 1) and to (b) AusE musicians (Experiment 2), it seems that although extensive experience with pitch either in the linguistic or in the musical domain allows learners to perceive lexical tone more accurately, distributional learning of lexical tones is only facilitated by domain-specific pitch experience. In fact, extensive pitch experience in a different domain (music, in this case) may actually interfere with distributional learning. We suggest that the lack of distributional learning in musicians may be partly due to an activation of a mode of processing that does not facilitate learning lexical tones from the distributional structure of the input. To be specific, AusE musicians in this study may have approached the task in a musical manner and relied on their existing robust musical representations to perceive the isolated lexical tones. Further studies are required to confirm our speculation. Nevertheless, the present experiment has demonstrated that distributional learning is modulated by domain-specific experience and that accurate encoding of the input does not necessarily predict success in distributional learning.

Acknowledgments

JHO was supported by the MARCS Institute via the Research Training Scheme. Portions of this work have been presented at the 2015 International Congress of Phonetic Sciences and the 2015 Society of Music Perception and Cognition Conference. We are grateful to Prof. Marcus Taft for his assistance with recruiting participants, to the MARCS Institute Music Cognition and Action group for their insightful comments on a previous draft, and to all the participants who volunteered their time to participate in this research. We are also grateful to Dr. Bharath Chandrasekaran and two anonymous reviewers for their comments.

References

- Alexander, J. A., Bradlow, A. R., Ashley, R. D., & Wong, P. C. M. (2011). *Music-melody perception in tone-language and non-tone-language speakers*. Poster presentation at the 156th Meeting of the Acoustical Society of America, Miami, FL.
- Alexander, J. A., Wong, P. C. M., & Bradlow, A. R. (2005). Lexical tone perception in musicians and non-musicians. In *Interspeech 2005* (pp. 397–400). Lisbon, Portugal: ISCA Archive.
- Baudoin-Chial, S. (1986). Hemispheric lateralization of Modern Standard Chinese tone processing. *Journal of Neurolinguistics*, 2, 189–199.
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). Timonium, MD: York Press.
- Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13–34). Amsterdam, the Netherlands: John Benjamins.
- Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain and Cognition*, 77, 1–10. doi:10.1016/j.bandc.2011.07.006

- 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. doi:10.1371/journal.pone.0060676
- Boersma, P., & Weenink, D.** (2013). *Praat: Doing phonetics by computer*. Retrieved from <http://www.praat.org>
- Burnham, D., Brooker, R., & Reid, A.** (2014). The effects of absolute pitch ability and musical training on lexical tone perception. *Psychology of Music*, 43, 881–897. doi:10.1177/0305735614546359
- Burnham, D., Ciocca, V., & Stokes, S.** (2001). *Auditory-visual perception of lexical tone*. Paper presented at EUROSPEECH 2001 Scandinavia, 7th European Conference on Speech Communication and Technology, Aalborg, Denmark.
- Burnham, D., Kasisopa, B., Reid, A., Luksaneeyanawin, S., Lacerda, F., Attina, V., ... Webster, D.** (2015). Universality and language-specific experience in the perception of lexical tone and pitch. *Applied Psycholinguistics*, 36, 1459–1491. doi:10.1017/S0142716414000496
- Chandrasekaran, B., Krishnan, A., & Gandour, J. T.** (2007). Mismatch negativity to pitch contours is influenced by language experience. *Brain Research*, 1128, 148–156. doi:10.1016/j.brainres.2006.10.064
- 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. doi:10.1016/j.bandl.2008.02.001
- Chao, Y.-R.** (1930). A system of tone-letters. *Le Maître Phonétique*, 45, 24–27.
- Chládková, K., Escudero, P., & Lipski, S. C.** (2013). Pre-attentive sensitivity to vowel duration reveals native phonology and predicts learning of second-language sounds. *Brain and Language*, 126, 243–252. doi:10.1016/j.bandl.2013.05.020
- Chobert, J., François, C., Velay, J.-L., & Besson, M.** (2014). Twelve months of active musical training in 8- to 10-year-old children enhances the preattentive processing of syllabic duration and voice onset time. *Cerebral Cortex*, 24, 956–967. doi:10.1093/cercor/bhs377
- Deutsch, D., Henthorn, T., & Dolson, M.** (2004). Absolute pitch, speech, and tone language: Some experiments and a proposed framework. *Music Perception*, 21, 339–356.
- Deutsch, D., Henthorn, T., Marvin, E., & Xu, H.** (2006). Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period. *The Journal of the Acoustical Society of America*, 119, 719–722. doi:10.1121/1.2151799
- Escudero, P.** (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. (Doctoral dissertation, Utrecht University, Utrecht, the Netherlands).
- Escudero, P.** (2009). The linguistic perception of similar L2 sounds. In P. Boersma & S. Hamann (Eds.), *Phonology in perception* (pp. 151–190). Berlin, Germany: Mouton de Gruyter.
- Escudero, P., Benders, T., & Wanrooij, K.** (2011). Enhanced bimodal distributions facilitate the learning of second language vowels. *The Journal of the Acoustical Society of America*, 130, EL206–EL212. doi:10.1121/1.3629144
- Escudero, P., & Williams, D.** (2014). Distributional learning has immediate and long-lasting effects. *Cognition*, 133, 408–413. doi:10.1016/j.cognition.2014.07.002
- François, C., Chobert, J., Besson, M., & Schön, D.** (2013). Music training for the development of speech segmentation. *Cerebral Cortex*, 23, 2038–2043. doi:10.1093/cercor/bhs180
- François, C., & Schön, D.** (2013). Neural sensitivity to statistical regularities as a fundamental biological process that underlies auditory learning: The role of musical practice. *Hearing Research*, 308, 122–128. doi:10.1016/j.heares.2013.08.018
- Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H.** (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences*, 19, 117–125. doi:10.1016/j.tics.2014.12.010
- Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M.** (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, 5, 171–189.
- Galle, M. E., & McMurray, B.** (2014). The development of voicing categories: A quantitative review of over 40 years of infant speech perception research. *Psychonomic Bulletin & Review*, 21, 884–906. doi:10.3758/s13423-013-0569-y
- Gomes, H., Molholm, S., Ritter, W., Kurtzberg, D., Cowan, N., & Vaughan, H. G., Jr.** (2000). Mismatch negativity in children and adults, and effects of an attended task. *Psychophysiology*, 37, 807–816. doi:10.1111/1469-8986.3760807
- Goudbeek, M., Cutler, A., & Smits, R.** (2008). Supervised and unsupervised learning of multidimensionally varying non-native speech categories. *Speech Communication*, 50, 109–125. doi:10.1016/j.specom.2007.07.003
- Gulian, M., Escudero, P., & Boersma, P.** (2007). *Supervision hampers distributional learning of vowel contrasts*. Paper presented at the International Congress of Phonetic Sciences, Saarbrücken, Germany.
- Krishnan, A., Swaminathan, J., & Gandour, J. T.** (2009). Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *Journal of Cognitive Neuroscience*, 21, 1092–1105. doi:10.1162/jocn.2009.21077
- Lee, C.-Y., Lekich, A., & Zhang, Y.** (2014). Perception of pitch height in lexical and musical tones by English-speaking musicians and nonmusicians. *The Journal of the Acoustical Society of America*, 135, 1607–1615. doi:10.1121/1.4864473
- 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. doi:10.1162/jocn.2010.21585
- Maye, J., & Gerken, L.** (2000). Learning phonemes without minimal pairs. In *Proceedings of the 24th Annual Boston University Conference on Language Development* (Vol. 2, pp. 522–533). Somerville, MA: Cascadilla Press.
- Maye, J., & Gerken, L.** (2001). Learning phonemes: How far can the input take us? In *Proceedings of the 25th Annual Boston University Conference on Language Development* (pp. 480–490). Somerville, MA: Cascadilla Press.
- Maye, J., Weiss, D. J., & Aslin, R. N.** (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, 11, 122–134.
- Maye, J., Werker, J. F., & Gerken, L.** (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82, B101–B111.
- McMurray, B., & Aslin, R. N.** (2005). Infants are sensitive to within-category variation in speech perception. *Cognition*, 95, B15–B26. doi:10.1016/j.cognition.2004.07.005
- Ong, J. H., Burnham, D., & Escudero, P.** (2015a). Distributional learning of lexical tones: A comparison of attended vs. unattended listening. *PLoS One*, 10, e0133446. doi:10.1371/journal.pone.0133446
- Ong, J. H., Burnham, D., & Escudero, P.** (2015b). Mandarin listeners can learn non-native lexical tones through distributional

- learning. In *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, Scotland: International Phonetic Association.
- Pajak, B., & Levy, R.** (2014). The role of abstraction in non-native speech perception. *Journal of Phonetics*, 46, 147–160. doi:10.1016/j.wocn.2014.07.001
- Paraskevopoulos, E., Kuchenbuch, A., Herholz, S. C., & Pantev, C.** (2012). Statistical learning effects in musicians and non-musicians: An MEG study. *Neuropsychologia*, 50, 341–349. doi:10.1016/j.neuropsychologia.2011.12.007
- Perfors, A., & Ong, J. H.** (2012). Musicians are better at learning non-native sound contrasts even in non-tonal languages. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 839–844). Austin, TX: Cognitive Science Society.
- Pfordresher, P. Q., & Brown, S.** (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attention, Perception, & Psychophysics*, 71, 1385–1398. doi:10.3758/APP.71.6.1385
- Poepel, T. J., & Weiss, D. J.** (2016). The influence of bilingualism on statistical word learning. *Cognition*, 152, 9–19. doi:10.1016/j.cognition.2016.03.001
- Qin, Z., & Mok, P.** (2012). The perception of speech and non-speech tones by tone and non-tone language listeners. In Q. Ma, H. Ding, & D. Hirst (Eds.), *Speech Prosody 2012* (pp. 366–369). Shanghai, China: ISCA Archive.
- Reeder, P. A., Newport, E. L., & Aslin, R. N.** (2013). From shared contexts to syntactic categories: The role of distributional information in learning linguistic form-classes. *Cognitive Psychology*, 66, 30–54. doi:10.1016/j.cogpsych.2012.09.001
- Reid, A., Burnham, D., Kasisopa, B., Reilly, R. G., Attina, V., Rattanasone, N. X., & Best, C. T.** (2015). Perceptual assimilation of lexical tone: The roles of language experience and visual information. *Attention, Perception, & Psychophysics*, 77, 571–591. doi:10.3758/s13414-014-0791-3
- Sadakata, M., Van der Zanden, L., & Sekiyama, K.** (2010). Influence of musical training on perception of L2 speech. In *Proceedings of the 11th Annual Conference of the International Speech Communication Association*, Makuhari (pp. 118–121). Makuhari, Japan: ISCA.
- Saffran, J. R., Aslin, R. N., & Newport, E. L.** (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928. doi:10.1126/science.274.5294.1926
- Schellenberg, E. G., & Trehub, S. E.** (2003). Good pitch memory is widespread. *Psychological Science*, 14, 262–266.
- Schön, D., & François, C.** (2011). Musical expertise and statistical learning of musical and linguistic structures. *Frontiers in Psychology*, 2, 167. doi:10.3389/fpsyg.2011.00167
- Shafer, V. L., Morr, M. L., Datta, H., Kurtzberg, D., & Schwartz, R. G.** (2005). Neurophysiological indexes of speech processing deficits in children with specific language impairment. *Journal of Cognitive Neuroscience*, 17, 1168–1180. doi:10.1162/0898929054475217
- Shook, A., Marian, V., Bartolotti, J., & Schroeder, S. R.** (2013). Musical experience influences statistical learning of a novel language. *The American Journal of Psychology*, 126, 95–104. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/23505962>
- Smith, L., & Yu, C.** (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, 106, 1558–1568.
- Stevens, C. J., Keller, P. E., & Tyler, M. D.** (2011). Tonal language background and detecting pitch contour in spoken and musical items. *Psychology of Music*, 41, 59–74. doi:10.1177/0305735611415749
- Terry, J., Ong, J. H., & Escudero, P.** (2015). Passive distributional learning of non-native vowel contrasts does not work for all listeners. In *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*. Glasgow, Scotland: International Phonetic Association.
- Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schröger, E.** (2005). Pitch discrimination accuracy in musicians vs non-musicians: An event-related potential and behavioral study. *Experimental Brain Research*, 161, 1–10.
- Van Lancker, D.** (1980). Cerebral lateralization of pitch cues in the linguistic signal. *Papers in Linguistics: International Journal of Human Communication*, 13, 201–277. doi:10.1080/08351818009370498
- Van Lancker, D., & Fromkin, V. A.** (1973). Hemispheric specialization for pitch and “tone”: Evidence from Thai. *Journal of Phonetics*, 1, 101–109.
- Van Lancker, D., & Fromkin, V. A.** (1978). Cerebral dominance for pitch contrasts in tone language speakers and in musically untrained and trained English speakers. *Journal of Phonetics*, 6, 19–23.
- van Leussen, J.-W., & Escudero, P.** (2015). Learning to perceive and recognize a second language: The L2LP model revised. *Frontiers in Psychology*, 6, 1000. doi:10.3389/fpsyg.2015.01000
- Wang, T., & Saffran, J. R.** (2014). Statistical learning of a tonal language: The influence of bilingualism and previous linguistic experience. *Frontiers in Psychology*, 5, 953. doi:10.3389/fpsyg.2014.00953
- Wang, Y., Jongman, A., & Sereno, J. A.** (2001). Dichotic perception of Mandarin tones by Chinese and American listeners. *Brain and Language*, 78, 332–348. doi:10.1006/brin.2001.2474
- Wanrooij, K., Boersma, P., & van Zuijen, T. L.** (2014a). Distributional vowel training is less effective for adults than for infants. A study using the mismatch response. *PLoS One*, 9, e109806. doi:10.1371/journal.pone.0109806
- Wanrooij, K., Boersma, P., & van Zuijen, T. L.** (2014b). Fast phonetic learning occurs already in 2-to-3-month old infants: An ERP study. *Frontiers in Psychology*, 5, 77. doi:10.3389/fpsyg.2014.00077
- Wanrooij, K., Escudero, P., & Raijmakers, M. E. J.** (2013). What do listeners learn from exposure to a vowel distribution? An analysis of listening strategies in distributional learning. *Journal of Phonetics*, 41, 307–319. doi:10.1016/j.wocn.2013.03.005
- Wayland, R. P., & Guion, S. G.** (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning*, 54, 681–712. doi:10.1111/j.1467-9222.2004.00283.x
- Werker, J. F., Yeung, H. H., & Yoshida, K. A.** (2012). How do infants become experts at native-speech perception? *Current Directions in Psychological Science*, 21, 221–226. doi:10.1177/0963721412449459
- Wilson, S. J., Abbott, D. F., Lusher, D., Gentle, E. C., & Jackson, G. D.** (2011). Finding your voice: A singing lesson from functional imaging. *Human Brain Mapping*, 32, 2115–2130. doi:10.1002/hbm.21173
- Wong, P. C. M., & Perrachione, T. K.** (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, 28, 565–585. doi:10.1017/S0142716407070312
- Yoshida, K. A., Pons, F., Maye, J., & Werker, J. F.** (2010). Distributional phonetic learning at 10 months of age. *Infancy*, 15, 420–433. doi:10.1111/j.1532-7078.2009.00024.x

Copyright of Journal of Speech, Language & Hearing Research is the property of American Speech-Language-Hearing Association and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.