Studies in Second Language Acquisition, 2015, **37**, 335–357.
doi:10.1017/S0272263114000849

FACTORS INFLUENCING SENSITIVITY TO LEXICAL TONE IN AN ARTIFICIAL LANGUAGE

Implications for Second Language Learning

Catherine L. Caldwell-Harris
Boston University

Alia Lancaster
University of Maryland

D. Robert Ladd
The University of Edinburgh

Dan Dediu
Max Planck Institute for Psycholinguistics

Morten H. Christiansen
Cornell University

This study examined whether musical training, ethnicity, and experience with a natural tone language influenced sensitivity to tone while listening to an artificial tone language. The language was designed with three tones, modeled after level-tone African languages. Participants listened to a 15-min random concatenation of six 3-syllable words. Sensitivity to tone was assessed using minimal pairs differing

Correspondence concerning this article should be addressed to Catherine L. Caldwell-Harris, Department of Psychological and Brain Sciences, Boston University, Boston, MA, 02215. E-mail: charris@bu.edu

© Cambridge University Press 2015

https://doi.org/10.1017/S0272263114000849 Published online by Cambridge University Press

only in one syllable (nonword task: e.g., *to-kà-su* compared to *ca-fí-to*) or only in tone (tone task: e.g., *to-kà-su* compared to *to-ká-su*). Proficiency in an East Asian heritage language was the strongest predictor of success on the tone task. Asians without tone language experience were no better than other ethnic groups. We conclude by considering implications for research on second language learning, especially as approached through artificial language learning.

Tone languages are notoriously difficult to learn as a foreign language for learners with no prior experience with a tone language (Francis, Ciocca, Ma, & Fenn, 2008; So & Best, 2011; Y. Wang, Spence, Jongman, & Sereno, 1999). As in all aspects of second and foreign language learning, there is considerable variability across individuals who try to learn a tone language after childhood (e.g., Lee, Vakoch, & Wurm, 1996). To better understand individual differences in sensitivity to linguistic tone in adults, we utilized a statistical learning paradigm (Saffran, Aslin, & Newport, 1996). Sensitivity to linguistic tone was quantified by assessing the ability to identify which three-syllable sequences were words in a miniature artificial language, in cases in which tone was the only feature that distinguished words from nonwords.

BACKGROUND

What Is Tone?

Voice pitch is used in different ways in different languages, and the most conspicuous typological difference is between tone languages and nontonal languages. Tone languages include pitch phonemes in addition to consonants and vowels, and pitch differences are used to distinguish one lexical item from another. In nontonal languages, pitch is used only for sentence-level distinctions of intonation, such as distinguishing questions from statements; these are sometimes referred to as *intonation languages*. Nontonal languages include English and most of the other languages of Europe. Standard examples of tone languages are Chinese, Thai, and Yoruba. The Yoruba words in (1) illustrate what it means to use pitch phonemes for lexical distinctions. The consonants and vowels are identical, and the words differ only in which pitch phonemes occur on the two syllables:[1]

(1) a. *ī̀gbā* [mid-mid] “two hundred”
b. *ī̀gbá* [mid-high] “calabash”
c. *ìgbà* [low-low] “time”
d. *ìgbá* [low-high] “(type of tree)”

https://doi.org/10.1017/S0272263114000849 Published online by Cambridge University Press

Although the difference between using and not using pitch for lexical distinctions seems clear enough, there are nevertheless intermediate cases for which the term *pitch accent language* is sometimes used. Relevant for our concerns in this article is the fact that most varieties of Japanese and some varieties of Korean (though not standard Seoul Korean) are usually classed as pitch accent languages (Hyman, 2009; Yip, 2002). Japanese and Korean *pitch accent* is in some ways comparable to English word stress (e.g., the difference between *forebear* "ancestor" and *forbear* "refrain from"), but it differs in two crucial respects. First, the prominence of the accented syllable in Japanese or Korean is not, as in English, a complex amalgam of pitch change, force of articulation, duration, and so on, but is purely a matter of pitch. For example, in Tokyo Japanese, the pitch accent is always manifested as a step down in pitch from the accented syllable to the following syllable, with no other accompanying phonetic differences. Second and more fundamentally, pitch accent is not obligatory: In any English word of two or more syllables, one syllable must be more prominent than the other(s), whereas in Japanese or Korean, it is possible to have words in which no syllable has an accent. The set in (2) from standard (Tokyo) Japanese illustrates this:

(2) a. *hasi desu* "it's a chopstick" (step down in pitch from *ha-* to *-si*; *ha-* is accented)
b. *hasi desu* "it's a bridge" (step down in pitch from *-si* to *desu*; *-si* is accented)
c. *hasi desu* "it's an edge" (no step down anywhere; neither syllable is accented)

It is clear from these examples that this kind of pitch accent language uses pitch differently from the way it is used in English, but at the same time it should be clear that pitch accent of this kind is also unlike tone in Yoruba, in which each syllable has its own pitch specification, and the specification is not a matter of the relation between one syllable and another.

In short, although there is no generally agreed on typology and no generally agreed on definition of pitch accent, linguists agree that Japanese and Korean use pitch in quite a different way from many other languages of East and Southeast Asia, such as Thai, Vietnamese, and Chinese languages such as Mandarin and Cantonese. In what follows we use the terms *tone language*, *pitch accent language*, and *intonation language* for convenience, but the reader should be aware that tricky conceptual issues are involved.

Factors Influencing Tone Sensitivity in a Second Language

Language Experience. Perceptual reorganization of linguistic tone in infancy is similar to changes that occur for vowel and consonant phonemes;

https://doi.org/10.1017/S0272263114000849 Published online by Cambridge University Press

this takes place between 4 and 9 months of age (Mattock & Burnham, 2006; Mattock, Molnar, Polka, & Durnham, 2007). But processes involved in tone language acquisition in adulthood remain poorly understood (So & Best, 2011; Y. Wang et al., 1999). Of potential relevance to second language learning are studies in which native Mandarin speakers were better able to discriminate unknown Cantonese tone words than were English speakers, demonstrating the advantageous effects of tonal language experience (Lee et al., 1996). Similarly, Wayland and Guion (2004) found that, in a discrimination task of Thai tone words, native Cantonese speakers outperformed native English speakers before and after a training session (see also Wayland & Li, 2008).

So and Best (2011) investigated the ability to learn to discriminate words in a tone language, Mandarin, of three groups: native speakers of Cantonese (also a tone language), native speakers of Japanese (a pitch accent language), and native speakers of English (an intonation language). As expected, English speakers had the most difficulty in this task. If prior tone language experience is crucial, then the Cantonese speakers should have outperformed the Japanese speakers. Unexpectedly, the Cantonese and Japanese speakers were similar in this task, indicating that the pitch features of Japanese were just as helpful for learning Mandarin tones as were Cantonese tones. This study suggests that experience with a language in which pitch features have a lexical function provides an advantage in dealing with an unfamiliar tone language.

In addition to language experience, certain types of tone instruction may be more beneficial to a learner with a nontonal native language. The gestural representations of Mandarin pitch contours facilitated tone discrimination and retention of semantic meaning in novice learners during an experimental session (Morett & Chang, 2014), and an alphabetic script plus a graphical representation of the pitch contours facilitated tone discrimination in learners in the classroom over the course of a semester (Liu et al., 2011). A study on individual differences demonstrated that participants with better abilities in pitch contour identification were less affected by high variability during vocabulary learning (Perrachione, Lee, Ha, & Wong, 2011).

Musical Ability. Linguistic tone experience and genetic heritage may facilitate musical training. There is a large literature concerning the phenomenon of absolute pitch, or the ability to name or produce a particular pitch in the absence of a reference note. The prevalence of absolute pitch is higher in Asian populations, even controlling for age of onset of musical training (Deutsch, Henthorn, & Dolson, 2004; Deutsch, Henthorn, Marvin, & Xu, 2006). There is also evidence that musical ability and achievement differ across ethnic groups, which could indicate the influence of genetic effects. In a task involving the recognition of pitch intervals, Hove, Sutherland, and Krumhansl (2010) found that Chinese

https://doi.org/10.1017/S0272263114000849 Published online by Cambridge University Press

and Korean participants were more accurate than Caucasians, with no difference between Chinese and Korean speakers in performance. Whereas Chinese and Korean populations can both be considered ethnically East Asian, Korean (as noted earlier) is not a tone language, though some dialects are generally regarded as pitch accent languages. This suggests that East Asian ethnicity, even for speakers without tone language experience, may affect musical ability.

Furthermore, musical training appears to facilitate the learning of linguistic tone. Native English speakers with no previous exposure to a tone language were trained on English pseudowords with pitch patterns resembling three Mandarin tones; each word was paired with a drawing to identify word meaning (Wong & Perrachione, 2007). Crucial stimuli to learn were minimal pairs based on tone. The majority (seven out of eight) of successful learners were amateur musicians. Among the less successful learners, only one out of nine was an amateur musician. Alexander, Wong, and Bradlow (2005) also found that American English–speaking musicians were more accurate in discriminating monosyllabic words with Mandarin tones than American English-speaking nonmusicians. Therefore, it appears that for people with no tone language experience, musical training aids in linguistic tone perception.

Genetics. Tone languages cluster in specific geographical regions. The indigenous languages of sub-Saharan Africa and East and Southeast Asia are generally tone languages, whereas those of Europe, North Africa, and South and Central Asia are not. Particularly in view of the apparent ethnic effects on musical pitch abilities, this clustering has raised the question of whether genetic biases could influence the learning of linguistic tone. Dediu and Ladd (2007) examined the population frequency of different alleles of two genes, *ASPM* and *Microcephalin*, which are involved in brain growth and development. They reported a negative correlation between the population frequency of the relatively recent so-called derived alleles of the two genes and the global geographical distribution of tone languages. Populations in which languages are more likely to be tonal have lower frequencies of the derived alleles of both genes, in particular, *ASPM*. Dediu and Ladd proposed that genetic differences may induce an individual cognitive bias that would favor or disfavor using pitch in lexical representations. Over long periods of intergenerational transmission, this bias would cause languages to drift in one direction or the other.

Using a battery of linguistic tone- and pitch-perception tasks and functional magnetic resonance imaging (fMRI) on genotyped participants, Wong, Chandrasekaran, and Zheng (2012) tested Dediu and Ladd's (2007) hypothesis. All participants were English-speaking adults with no experience with a tone language. A positive correlation was obtained between tone identification and individuals' genetic load of the derived alleles of *ASPM*.

https://doi.org/10.1017/S0272263114000849 Published online by Cambridge University Press

No effect was found for *Microcephalin*, and no correlations were found between individual *ASPM* makeup and a task involving the manipulation of segmental (nonpitch) phonemes. This suggested that the derived *ASPM* alleles are relevant to the perception or encoding of linguistic tone, in line with Dediu and Ladd's hypothesis. Wong et al. (2012) actually interpret their findings as showing that the derived alleles favor acquisition of tone, which is the opposite of what Dediu and Ladd suggested. Space does not permit us to discuss this issue at any length, but the essence of an alternative explanation for Wong et al.'s results runs as follows: What their experiments tested was the not the ability of individuals to process or acquire tone but rather their ability to separate pitch from the rest of the speech signal—that is, to abstract it away from the consonants and vowels and to make metalinguistic judgments about it. Such separation is precisely the opposite of what is called for in processing a tone language, in which the pitch and the vowel are closely integrated in the lexical identity of a given word, but is more or less required in an intonation language, in which lexical identity depends on separating out the pitch contour and ignoring it. On this interpretation, the effect of the derived alleles would have been to favor the separation of pitch from the rest of the signal and hence to disfavor the inclusion of pitch specifications in lexical representation.

We have not genotyped research participants in the current study but instead are exploring predictions of the genetic account by including individuals with East Asian and African heritage. Genetic influences on lexical tone will be supported if persons of East Asian or African ethnicity show more sensitivity to lexical tone than persons of other ethnicities, even if they grew up with little or no fluency in a tone language.

Statistical Learning and Artificial Language Learning

The work reported here was an exploratory study based on the evidence just reviewed and was designed to find out whether certain factors (i.e., ethnic background, experience with a tone language, and musical ability) independently account for above-average sensitivity to tones presented as part of an artificial language. The artificial language paradigm has been used to demonstrate that infants and adults use the statistical distribution of cues in artificial languages to segment words via transitional probabilities (e.g., Saffran, Newport, & Aslin, 1996). These results have led to extensive testing of hypotheses about what strategies and knowledge infants and adults use when extracting regularities from a novel stream of sounds.

https://doi.org/10.1017/S0272263114000849 Published online by Cambridge University Press

Adult speakers have been shown to be able to segment words based on nonadjacent statistical dependencies (e.g., Peña, Bonatti, Nespor, & Mehler, 2002). In Peña et al.'s (2002) study, adult French speakers were given a continuous stream of an artificial language with three-syllable words that had nonadjacent transitional probability. For instance, syllables one and three often appeared in the same place in many words, in an AXB fashion, making the transitional probability between A and B high when they are not adjacent to each other. It was found that the participants used this cue to segment the artificial language into words. Onnis, Monaghan, Richmond, and Chater (2005) demonstrated the same results as Peña and colleagues with English speakers. Onnis et al. also determined that nonadjacent dependencies for use in segmentation are subject to phonological properties of the syllables within the words and by the phonotactic constraints of the native language.

Saffran and colleagues (Saffran, Johnson, Aslin, & Newport, 1999) applied their paradigm to the segmenting of nonlinguistic (musiclike) sequences of tones. Their stimulus stream was similar to that of Saffran, Newport, et al. (1996) with regard to the transitional probability of the three-syllable "words," but each syllable was replaced with a distinct tone (e.g., *pa* was replaced by D-sharp). Both adults and infants were able to successfully segment the tone stream into "words" at levels better than chance, and adults were significantly better at segmenting words with higher transitional probabilities.

A recent study by T. Wang and Saffran (2014) aimed to test statistical learning of an artificial language designed with tonal properties in mind. Their language consisted of three trisyllabic words, constructed from a total of nine unique consonant-vowel syllables that differed both segmentally and tonally in their phoneme composition (i.e., there were nine different consonant-vowel sequences and nine different tonal patterns, and each tonal pattern was uniquely matched with a consonant-vowel sequence). Because tonal regularities were therefore redundant, it was possible to segment the language without paying attention to the tonal properties. Indeed, testing involved distinctions in syllable ordering only between words and nonwords (in which the order of syllables was reversed).[2] In contrast to natural languages in which tone is typically used to signal meaning differences for minimal sets of the same segmental sequence (as in [1]), words in T. Wang and Saffran's artificial language can be distinguished purely in terms of their different segments. Thus, rather than investigating the role of tone in artificial language learning, their study may be better thought of as testing statistical learning in the context of "exotic"-sounding stimuli, which may explain why bilinguals outperformed monolinguals in their experiments. It is therefore still an open question whether adults can learn to integrate tone into their representations of lexical candidates in a statistical learning task. This is the question that motivated our study.[3]

https://doi.org/10.1017/S0272263114000849 Published online by Cambridge University Press

THE CURRENT STUDY

The present study used the statistical learning paradigm to test individuals' sensitivity to tone as a cue to identifying the "words" (recurring sequences of syllables) in an artificial language. Adults listened to 15 min of an artificial language consisting of three-syllable sequences in which the middle syllable varied in tone. After this initial exposure, participants were tested to see whether they could distinguish sequences of syllables that had occurred in the artificial language from sequences that had not. In what we refer to as the nonword task, participants encountered correct and incorrect test sequences of syllables that differed in the segmental makeup of the syllables, similar to the work by Saffran and colleagues (e.g., *fu-sì-ke* and *po-sí-cu*). The crucial additional test for measuring sensitivity to lexical tone was what we refer to as the tone task: Here the incorrect three-syllable sequences differed only in tone from the segmentally identical correct sequences (e.g., *fu-sì-ke* and *fu-sí-ke)*.

Designed as an exploratory study, this research fills two gaps in the current literature. First, it examines behavior within a controlled environment in which listeners need to extract regularities, including those related to lexical tone, from a continuous speech stream, as occurs during normal language acquisition. Other studies that have measured tone sensitivity have used isolated vocabulary words during training before a discrimination task (e.g., Wong & Perrachione, 2007). Second, it assesses in a single task how language experience, ethnicity, and musical training influence sensitivity to lexical tone language learning, unlike other studies that measured one or two factors per task. Thus, the study may have implications for the potential impact of first language experience on the acquisition of a second language (here modeled via artificial language learning). Moreover, to the best of our knowledge, this is the first study to apply a statistical learning paradigm in the context of an artificial tone language, in which minimal pairs of words differ in their tonal properties (as in natural tone languages).

METHOD

Participants

A total of 150 undergraduates (104 female; M_{age} = 19.34, SD = 1.69) from Boston University participated in the current study, for 3 hr of psychology course credit. A recruitment flyer offered Psychology 101 credit for anyone who was interested. Two related flyers additionally recruited Psychology 101 students who identified as having African or East Asian heritage. According to the self-reported ethnicity of the participants,

https://doi.org/10.1017/S0272263114000849 Published online by Cambridge University Press

the resulting sample contained 45 people of European descent (Caucasians, including eight Hispanics; 91% native English speakers), 14 African Americans (including immigrants from Caribbean and African countries; 100% native English speakers), 83 people with an East or Southeast Asian background (61% native English speakers), and eight persons who labeled their background as Middle Eastern or Indian (75% native English speakers). For simplicity, we refer to those with East or Southeast Asian background simply as Asian and note that this does not include Middle Eastern and Indian. The members of the Asian group were subdivided into three subgroups on the basis of their exposure to tone languages, as described in more detail in the next paragraph.

Participants completed an extensive language questionnaire that asked them to list languages with which they had some experience and to rate their proficiency on a scale from 1 to 7 (7 = *native speaker*, 6 = *near native*, 5 = *very good*, 4 = *good*, 3 = *fair*, 2 = *some*, 1 = *little*; in later analyses the value 0 was included to indicate no knowledge of an East Asian language). We recorded the number of languages that each participant had been exposed to, including those with self-ratings of 1. We also classified participants according to whether they had experience with a tone language. This was complicated by the difficulty of defining both *tone language* and *experience*. With regard to experience, many participants denied knowing the language of their parents or grandparents but said they knew some words or had some partial fluency. We therefore used a proficiency rating of 3 (*fair*) or higher as a cutoff. Those who rated their proficiency in a tone language as 0–2 (indicating no, little, or some proficiency) were classified as not having tone language experience. As for the definition of a tone language, we initially classified Japanese and Korean as nontonal languages and grouped those with experience in Japanese or Korean together with, for example, Chinese Americans who grew up with minimal exposure to their heritage language. However, inspection of raw data revealed, surprisingly, that those with experience in Japanese or Korean were just as accurate on the tone task as those with experience in tone languages. We therefore created a third group (Asian Japanese/Korean), which was distinct from those with and without Asian tone language experience. The three subgroups of Asian participants, therefore, were as follows: Asian tone language (those with a self-rating of 3 or higher in Thai, Vietnamese, or a Chinese language), Asian Japanese/Korean (those with a self-rating of 3 or higher in Japanese or Korean), and Asian none (all others reporting East or Southeast Asian ethnicity). The last group included a few participants with high self-ratings in Asian nontone languages, specifically Malay and Tagalog.

We also administered a music history questionnaire. Musical proficiency in each instrument, including voice, was rated on a scale of 0 (none) to 10 (extremely proficient/professional). For analysis, the highest rating a

https://doi.org/10.1017/S0272263114000849 Published online by Cambridge University Press

participant listed (for any instrument) was used as the measure of musical ability.

To determine if sensitivity to tone in the statistical learning paradigm could be explained by cognitive ability or cognitive style, we administered an IQ test and the Need for Cognition Questionnaire (NfC; Cacioppo, Petty, & Kao, 1982). The IQ test used was Scale 3, Form A of Cattell's Culture Free Intelligence Test (CFIT; Cattell, 1971). This test is nonlinguistic, consisting of four timed subsections of multiple choice problems pertaining to visual-spatial reasoning. It has previously been shown to correlate with proficiency in the second language learning of a miniature version of Russian (e.g., Kempe & Brooks, 2008; Kempe, Brooks, & Kharkhurin, 2010). The NfC (Cacioppo et al., 1982) measured participants' willingness to engage in cognitive activities. They rated their agreement or disagreement on a scale of –4 (*very strong disagreement*) to 4 (*very strong agreement*) on 34 statements such as "I prefer life to be filled with puzzles that I must solve."

Materials

The artificial language task consisted of three 5-min blocks of random combinations of six 3-syllable sequences ("words") such as *fu-sì-ke*, with the restriction that no word could occur twice in a row. The language was loosely modeled on African languages such as Yoruba, with three level tones: high, mid, and low; each pseudoword had a mid tone on the first and last syllables and either a high or a low tone on the middle syllable. This is quite unlike the tone patterns of most of the East and Southeast Asian languages with which some of our participants were familiar. Two versions of the artificial language were created, to allow sequences that were nonwords in one version to be the words in the other, and vice versa. As explained in the Procedure section, listeners were required to distinguish three-syllable sequences that had appeared during the listening period (e.g., *fu-sì-ke*) from those that had not (e.g., *po-sí-cu* or *fu-sí-ke*).

The stimuli were prepared by the third author. Each of the six intended test words (i.e., the three-syllable sequences) was embedded in several longer sequences of 15–20 consonant-vowel syllables written out in a broad phonetic transcription. This author then recorded all the sequences on a steady level pitch at a rate of about three syllables per second. From these recordings, a single token of each test word was chosen for use in the stimuli; tokens were chosen to be as uniform as possible in pitch and duration. The pitch of the middle syllable was then modified using Praat software (Boersma, 2001; www.praat.org), by raising or lowering the pitch of each analysis frame in the syllable by approximately

https://doi.org/10.1017/S0272263114000849 Published online by Cambridge University Press

3.5 semitones (i.e., between a minor and a major third) relative to the originally recorded pitch. The resulting words were then concatenated into the longer artificial language stimulus blocks.

The six words were *pacóti*, *sefípo*, *kitúca*, *fusìke*, *cupèfo*, and *tokàsu* (the first three words with mid-high-mid tonal sequence and the last three with mid-low-mid tonal sequence). The six nonwords were created using the same syllables in different combinations; they were *cafíto*, *ketúpa*, *sucóki*, *fokàse*, *posìcu*, and *tipèfu*. As with the words, the first three of these had a mid-high-mid tonal sequence, and the last three had a mid-low-mid tonal sequence. In the nonword task, each of the six words was paired with each of the nonwords to create 36 nonword trials. In half of the nonword trials, the word was played first in the two-alternative, forced-choice sequence, and in the other half, the nonword was played first.

In the tone task, the three-syllable segments were tone minimal pairs. For example, participants had to choose between *fu-sì-ke* (which appeared in the speech stream) and *fu-sí-ke* (which did not). The six tone minimal pairs were presented for evaluation four times each, twice with the correct answer played first in the sequence and twice with the incorrect variant appearing first, for a total of 24 tone-task trials. Crucially, correct performance on the tone task required the participants to integrate tonal information into their representation of the training stimuli.

Two versions of the artificial language were created to allow sequences that were nonwords in one version to be the words in the other and vice versa. A short example from the listening phase of the study as well as a minimal tone pair can be heard at www.lel.ed.ac.uk/~bob/SSLA-DEMO/.

Procedure

Each participant listened to the artificial language for three 5-min periods, presented via a computerized protocol implemented with E-Prime experimental control software. A short break was interposed between periods, and participants pressed a button when they were ready for the next period. Participants were not told they were listening to a language with individual words nor that the words contained lexical tone.

In the test phase following the listening phase, participants read instructions informing them that they would need to identify which words had occurred in the speech stream they had listened to. They heard two 3-syllable sequences and chose which had been present during the listening period. In the 36 nonword trials, participants heard pairs such as *fu-sì-ke* (which occurred in the speech stream) and *po-sí-cu* (which did not). In the 24 trials of the tone task, the three-syllable segments were tone minimal pairs. For example, participants had to choose between

https://doi.org/10.1017/S0272263114000849 Published online by Cambridge University Press

fu-sì-ke (which appeared in the speech stream) and *fu-sí-ke* (which did not). Stimuli from the tasks were blocked, with nonword trials appearing first, followed by tone trials. Blocking of the tasks was necessary because hearing a tone minimal pair was equivalent to being informed that its segments constituted a word. In addition, presenting the nonword (segmental) task meant that we could measure participants' sensitivity to the segmental properties of three-syllable words with them having been distracted by learning that tone was important for distinguishing sequences.

In a pilot version of the study with 25 participants, performance on the tone task was low (55% correct), and participants reported losing attention and/or feeling sleepy during the listening phase. To assure greater attention to the speech stream, in the final version of the study participants were told to monitor the speech stream and to respond with a button press when they heard repeated syllables (e.g., *ba ba*). Twenty repeated syllables were inserted into the language for this purpose. Thus, in the final version of the listening task, there was a one-third probability that the initial or final syllable in one of the six words in a block would be repeated. It is noteworthy that this cover task did not succeed in increasing accuracy on the tone task, as the final average on the tone task across the 150 participants was only 56% correct, suggesting that the cover task neither helped nor harmed participants' performance. Tone-task accuracy also did not vary with correctness on the cover task.

Participants completed a debriefing questionnaire that asked them whether they had realized when listening to the artificial language that the "words" in the stream had three syllable tones, whether tones were an important part of word identity, and whether they felt they were able to learn the words.

Finally, we note that participants in the study also completed two other tasks related to pitch perception in addition to the tasks we have reported here. One task involved the perception of tones with missing-fundamental frequency. It is known (e.g., Schneider et al., 2005) that there are individual differences in the way such tones are perceived, with some listeners perceiving the pitch to be at the level of the missing fundamental and others basing their pitch percept primarily on the partials that are actually present in the tone. We have included the results of this task in a longer paper (Ladd et al., 2013) on individual differences in missing-fundamental perception. For our present purposes it is enough to note that we found no connection between missing-fundamental perception and sensitivity to linguistic tone. The other additional task was a strictly exploratory study in which both the frequency and intensity of stimulus tones were manipulated. We were looking for individual differences in the extent to which changes in intensity across a pair of stimulus tones could mask the ability to detect frequency differences

https://doi.org/10.1017/S0272263114000849 Published online by Cambridge University Press

that would otherwise be clearly perceptible. It would take us too far afield to describe the task in detail, but it is worth noting that the performance of participants with experience in an East Asian language appears distinct from that of all other groups: Roughly speaking, they were more likely to detect small frequency differences even in the presence of potentially distracting intensity changes. At the very least, this finding is not incompatible with the results in the current paper.

RESULTS

Figure 1 plots the results of the tone and nonword tasks for the three groups of Asians and the non-Asian group. No group differences were obtained in the nonword task, $F(3, 146) = 0.85$, $p = .47$, and the accuracy of all groups was significantly above 50%, which was chance performance for the two-alternative, forced-choice task. On the tone task, however,

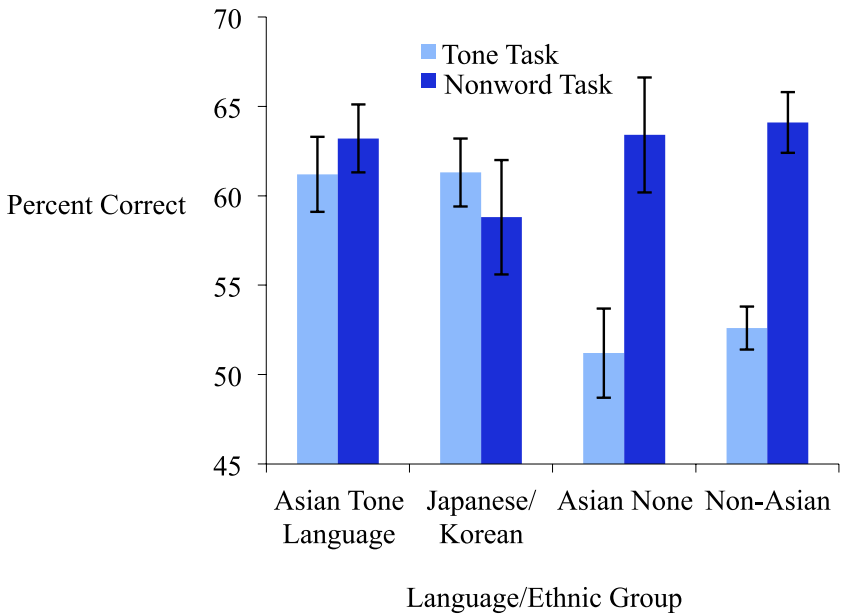


Figure 1. The percent correct on the tone task (distinguishing minimal pairs that differ by a tone change in one syllable) and nonword task (distinguishing minimal pairs that differ by a segment change in a syllable) for participants descended from non-East Asians and three groups of East Asians categorized by knowing an East Asian language at the level of *fair* or better. All bars represent group performance that is above chance with a probability of $p < .05$, except that of the Asian none group, which performed no differently from chance on the tone task. Error bars are the standard error of the mean.

https://doi.org/10.1017/S0272263114000849 Published online by Cambridge University Press

a one-way ANOVA revealed significant group differences, $F(3, 146) = 7.7$, $p < .001$. Tukey's LSD post hoc test was used to verify the significance of between-group differences. Asians with experience in either a tone language or Korean/Japanese were the most accurate on the tone task and did not differ from one another, as is apparent in Figure 1. We will thus refer to these Asians as having East Asian language experience. Those in the Asian none group did not differ significantly from non-Asians via post hoc comparison. However, the tone-task accuracy of the Asian none group did not differ from chance, whereas the accuracy of the non-Asian group was slightly above chance, $t(66) = 2.2$, $p = .34$.

Figure 1 summarizes our main empirical result but leaves open the question of why this pattern of differences occurred. To provide a basis for inferring the causes of this group difference, in Table 1 we list the group averages for the other variables studied (standard deviations appear in parentheses). Table 1 also provides an ethnic breakdown of the non-Asian group. Caucasians, as a group, had slightly above chance accuracy, $p = .02$, whereas Indians/Middle Easterners and African Americans performed at chance. However, the differences among these three groups of non-Asians were not statistically reliable, via post hoc comparisons. It was for this reason that we grouped the non-Asians together in Figure 1.

The CFIT and NfC scale did not vary across ethnic or language groups (see the group averages in Table 1). These test scores also did not correlate with performance on the tone task (CFIT, $r = .12$; NfC, $r = -.04$). The lack of correlation with group performance suggests that neither general cognitive ability (as measured by the CFIT) nor willingness to engage in the experimental task (measured by the NfC) appear to explain the group differences in the tone language task.

Bilingualism and East Asian Language Experience

Group differences did exist for two language-learning variables: best second language (i.e., best L2) and number of languages (see Table 1). This raises the possibility that group differences on the tone task, observable in Figure 1, may be due to bilingualism. On the language-learning questionnaire, participants self-rated their proficiency in all their languages. The best L2 variable was determined according to the language that received each participant's second-highest rating of proficiency. This variable ranged from 0 to 7, as some participants reported native-speaker ability in more than one language. In a one-way ANOVA, best L2 varied across the six groups, $F(5, 144) = 14.2$, $p < .001$. Post hoc comparisons revealed that high ability in a second language effectively divided the six groups into two groups of three, which cut across ethnicity.

https://doi.org/10.1017/S0272263114000849 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263114000849 Published online by Cambridge University Press

Table 1. Group averages for key variables

Group name (*n*)	Best L2 (0–7 scale)	Number of languages	Musical ability (0–10 scale)	NfC score	CFIT score	Accuracy nonword test (%)	Accuracy tone test (%)
Non-East Asians							
European (45)	3.7 (1.7)	2.7 (1.0)	3.1 (2.9)	40.0 (28.2)	25.6 (7.3)	**65 (0.1)**	**54 (0.1)**
African descent (14)	3.4 (1.3)	2.6 (0.9)	2.6 (2.4)	37.1 (21.9)	25.6 (6.0)	**64 (0.1)**	52 (0.1)
Indian or Middle Eastern (8)	5.8 (1.2)	3.5 (1.1)	2.8 (2.8)	46.0 (27.3)	26.1 (5.7)	**59 (12.0)**	46 (0.1)
East Asians							
Asian tone language (44)	5.8 (1.7)	3.3 (0.9)	4.5 (3.0)	36.3 (23.6)	37.6 (5.4)	**63 (0.1)**	**61 (0.1)**
Japanese/Korean (21)	5.9 (1.3)	3.3 (0.9)	4.8 (2.6)	24.8 (24.6)[a]	25.6 (8.8)	**59 (0.2)**	**61 (0.1)**
Asian none (18)	3.3 (2.2)	2.9 (1.2)	4.2 (2.8)	37.9 (28.1)	25.0 (10.0)	**63 (0.1)**	51 (0.1)
Statistical significance	***	*	*	n.s.	n.s.	n.s.	***

Note. L2 = second language, NfC = Need for Cognition Questionnaire, CFIT = Culture Free Intelligence Test, and n.s. = not significant. Above-chance performance is indicated by bolding. Standard deviations appear in parentheses.

[a] The low value is an average that includes three low-scoring Koreans, although note that no significant group differences were found.

*$p < .05$. ***$p < .001$.

One grouping consisted of Caucasians, African Americans, and the Asian none category. On average, these groups had not learned a second language beyond the level of *fair* (i.e., a rating of 3). These three groups did not differ among one another in their best L2 averages but did differ from the second grouping, whose members reported higher proficiency in a second language (all post hoc comparisons, $p < .01$). This group included many of the international students in our sample. They were ethnically from India or the Middle East or were proficient in an East Asian language.

What is noteworthy about the grouping revealed by best L2 is that bilingualism predicted tone-task performance for East Asians but not for the Indian/Middle Eastern group, whose members had high best L2 ratings. Their degree of bilingualism was similar to those with the highest tone-task performance, but the Indian group was at chance on the tone task.

Our other index of language-learning ability was a report of the number of languages (see Table 1). This showed a similar, although statistically weaker, pattern. A one-way ANOVA on this variable showed differences across the six groups, $F(5, 144) = 2.7$, $p = .02$. Post hoc tests showed that Indians listed familiarity with a larger number of languages than did African Americans and Caucasians, $p < .05$. Those in the Asian none group reported a lower number of languages than did the other Asians, although statistically this was only a trend, $p < .08$.

Looking at group averages, we can see that bilingualism did not facilitate tone-task accuracy, as the Indian group was highly bilingual but did poorly on the tone task. Having East Asian ethnicity also did not facilitate performance on the tone task, as the Asian none group had East Asian ethnicity but low tone-task accuracy. It seems that what was needed was both East Asian ethnicity and a high degree of bilingualism. Also sufficient, for this data set, was East Asian language experience. Note that in this data set, individuals who had East Asian language experience were identical to those who had high bilingualism and Asian ethnicity.

Multiple regression was used to explore these variables at the level of individual participants. As a prelude to multiple regression, raw correlations were calculated between participant variables and the two measures of statistical regularity extraction: the tone task (measured sensitivity to tone in the artificial speech stream) and the nonword task (measured sensitivity to segments in the speech stream).

Table 2 shows correlations among the two tasks, ethnicity, East Asian language proficiency (on a scale from 0–7), the participants' proficiency score in their best foreign language, and musical ability. Nonword-task accuracy did not correlate with any other variables (see the shaded cells). Tone-task accuracy was most strongly correlated, at $r = .36$, with East Asian language experience (which, as noted, in our sample was knowledge

https://doi.org/10.1017/S0272263114000849 Published online by Cambridge University Press

Table 2. Correlation matrix for variables that showed group differences

Variables	Tone task % correct	Nonword task % correct	Asian ethnicity (yes/no)	East Asian language proficiency	Best L2 proficiency
Nonword task	0.15				
Asian ethnicity	0.27	–0.07			
East Asian language	0.36	–0.11	0.74		
Best L2	0.25	–0.16	0.34	0.64	
Musical ability	0.17	0.03	0.23	0.17	0.09

Note. L2 = second language. $N = 150$. Proficiency self-report ratings are on a 1–7 scale. Shaded cells represent r values that are not significant at $p < .05$.

of the languages Mandarin, Cantonese, Thai, Vietnamese, Korean, or Japanese). Statistically significant (although weak) correlations were also obtained between task accuracy and Asian ethnicity, best L2, and musical ability. These correlations could be noncausal, deriving from their correlations with East Asian language ability. This was confirmed when the four predictors from Table 2 (East Asian ethnicity, East Asian language proficiency, best L2, and musical ability) were entered as predictors in an exploratory multiple regression on tone-task accuracy. Only East Asian language proficiency explained unique variance on the tone task; $\beta = 0.30$, $b = 0.01$, $t(145) = 2.1$, $p = .04$. The other three predictors were nonsignificant, with regression statistics as follows: Asian ethnicity, $\beta = 0.00$, $b = 0.00$, $t(145) = 0.02$, $p = .98$; bilingualism (best L2), $\beta = 0.05$, $b = 0.00$, $t(145) = 0.49$, $p = .62$; and musical ability, $\beta = 0.12$, $b = 0.00$, $t(145) = 1.4$, $p = .15$.

Individual Differences Exploration

To better understand the influence of language experience on tone-task accuracy, we examined the language experience and tone-task performance of non-Asians (recall that in our categorization non-Asians included Indians). We examined the performance of non-Asians who reported being exposed to a tone language (including Japanese and Korean) to see whether it matched the performance of East Asians with such language experience. Unfortunately, in our sample there were only four non-Asians with such experience at the self-reported level of 3 or higher on our 7-point scale: two African Americans (Japanese at Level 3 and Benin at Level 4), one Caucasian (Japanese at Level 4), and one Indian (Zulu at Level 3). Their average performance on the tone task was 47% (i.e., at chance).

https://doi.org/10.1017/S0272263114000849 Published online by Cambridge University Press

We also examined individual performance among the Asian group (again, recall that in our categorization this included only East and Southeast Asians). Asians who reported tone language proficiency in the fair and good range (3 and 4) averaged 56% on the tone task. Those with proficiencies in the range of 5–7 (indicating very good to native speaker abilities) had an accuracy of 63% correct. This suggests that greater East Asian language experience facilitated sensitivity to tone in the artificial-language learning task. Grouping together all Asians in our sample revealed that tone-task accuracy was a linear function of self-reported proficiency in a tone or pitch accent language, $r = .34$, $p < .001$. Indeed, the correlation across all 150 participants in Table 2 suggests that one could regard the Asian none group as part of the lower tail of the distribution of tone language proficiency, not as a separate group; its members were provisionally treated as a separate group as part of testing the genetic hypothesis.

Why did fair-to-good proficiency with Japanese, Benin, or Zulu fail to help the non-Asians in the tone task, or at least to move them up to the 56% average of the Asians with fair-to-good tone (or pitch accent) language proficiency? Although no conclusions can be drawn from a few individuals, an obvious possible explanation is that the Asians were all heritage learners of Mandarin, Cantonese, Thai, Korean, or Japanese and thus had their exposure in early childhood, from their parents in the home. The non-Asians had their exposure in middle childhood or the teen years, mostly from classroom foreign language learning. Early, naturalistic language acquisition may promote automatic sensitivity to the features of a sound stream, in a manner that is then helpful when extracting regularities while listening to an artificial language. This speculation is consistent with training studies: Persons who had childhood exposure to Hindi or Zulu were able to learn (or relearn) subtle phonemic contrasts in those languages, whereas matched controls with no prior exposure showed no such sensitivity (Bowers, Mattys, & Gage, 2009).

The lack of effect of musical experience on sensitivity to tone was also surprising (but see also T. Wang & Saffran, 2014, for a similar result). Previous research has revealed relationships both between musical abilities and tone language as well as between musical experience and artificial language learning (Chobert & Besson, 2013; Schön & François, 2011).

Musical Ability, Task Trade-Offs, and Awareness of Tone

Surprisingly, musical ability was only weakly correlated with tone-task performance, $r = .17$, $p = .04$, and was not significant in an exploratory multiple regression on tone task with musical ability and East Asian

https://doi.org/10.1017/S0272263114000849 Published online by Cambridge University Press

language experience as predictors. Musical ability did not facilitate tone-task accuracy for those without East Asian language experience, as that correlation was nonsignificant, $r = -.09$, $p = .40$.

Table 2 shows that the correlation between accuracy in the tone and nonword tasks was nonsignificant, when calculated across the entire sample of participants. This indicates the absence of task trade-offs. That is, sensitivity to tone did not detract from sensitivity to segmental changes nor vice versa. Examining correlations between the two tasks within the groups revealed a weakly positive correlation between the two tasks for those with East Asian language ability, $r = .29$, $p = .02$, but no correlation for other participants, $r = .10$, $p = .30$.

The debriefing questionnaire was analyzed to determine how much awareness listeners had about the task and whether awareness was related to scores on the tone task and nonword task. We were particularly interested in the question, "When listening to the artificial language, were you aware that the tone of the syllables mattered?" During the pilot study, the experimenters recorded anecdotes from an in-person oral interview. Some Caucasian participants reported being dumbfounded when the test trials appeared and when they were asked to judge which of two apparently identical sequences, *fu-sì-ke* or *fu-sí-ke*, had appeared in the speech stream. In contrast, the Asian participants who were fluent in a tone language appeared less surprised by the tone-task trials. Note that these anecdotes of being surprised (or not) emerged during the in-person social interaction that was part of the pilot study. However, responses to the written debriefing questionnaire showed only modest group differences on this question, although in the expected direction. The percentages of those who reported being aware that tone mattered were as follows: 57% of Asians with East Asian language experience, 50% of Asians without such experience, and 44% of Caucasians.

DISCUSSION

The ability to detect word boundaries in a stream of novel syllables in which tone distinguishes words from nonwords was superior (compared to other groups) for persons of East Asian heritage who reported having fair or better proficiency in a tone language or Japanese or Korean. Their accuracy rate of 61% is similar to the rate obtained in other artificial language tasks with adults (e.g., Saffran, Newport, et al., 1996). This confirms findings that artificial language learning is influenced by prior language experience (Bent, Bradlow, & Wright, 2006). The difficulty posed by our tone task for participants without tone or pitch accent proficiency is a novel finding in the literature on statistical learning paradigms.

One of the surprises in the current study was that knowledge of Japanese or Korean was just as helpful on the tone task as knowing one

https://doi.org/10.1017/S0272263114000849 Published online by Cambridge University Press

of the tone languages Cantonese, Mandarin, Thai, or Vietnamese. This is consistent with the finding of So and Best (2011) that knowledge of Japanese was just as helpful as knowledge of Cantonese for learning Mandarin tones. This may have implications for the typology of how pitch features are used in language: Japanese and Korean, despite being very different from Chinese or Vietnamese, appear to provide relevant language experience that enhances performance on our tone task.

East or Southeast Asian ancestry alone, without corresponding language experience, did not provide any advantage on the tone task. This appears to contradict findings in the literature that people of East or Southeast Asian ancestry have advantages in pitch-related tasks. However, some of the findings on the Asian advantage in pitch tasks may also reflect language experience, such as Hove et al.'s (2010) finding that both Koreans and Chinese speakers had an advantage in a relative-pitch task, compared to other groups.

IMPLICATIONS FOR SECOND AND FOREIGN LANGUAGE LEARNING AND FUTURE RESEARCH

Learning a tone language as a second or foreign language obviously requires paying attention to tone. This attention can be explicit or implicit (automatic). Automatic attention to features of a new language will be influenced by what features have been most useful in one's lifetime of language processing. That is, learning a new language—including a tonal language—is easier when it is more similar to a language the learner already knows. Instruction on lexical tone can make this attention more explicit (Liu et al., 2011; Morett & Chang, 2014). Instruction can also be tailored to take advantage of individual differences in tone perception (Perrachione et al., 2011).

In our artificial learning task, listeners were not warned ahead of time that tones were relevant for determining which sequences would be considered words in the artificial language. This suggests that those individuals who were successful in discriminating tone minimal pairs were those who automatically took into account the fact that tone varied in the speech stream. We found that the most important predictor of listeners' performance on this task was their degree of self-rated fluency in an East Asian language. This, in turn, suggests that listeners' prior experience with lexically meaningful pitch distinctions facilitates automatic attention to tone in the sound stream. Future work with artificial languages and training tasks can examine whether and how learning changes with instructions emphasizing ahead of time the importance of attending to tone in the speech stream. Other manipulations to encourage automatic attention to tone could include pretraining tonal sequences to have emotionally charged meanings. Still, an obvious implication of our study is to

https://doi.org/10.1017/S0272263114000849 Published online by Cambridge University Press

reinforce the commonsense understanding that speakers of nontonal languages learning a tone language must overcome a lifetime's habit of discounting tone as a contributor to lexical identity.

We used a naturalistic sample, meaning we tested everyone who wanted to be in the study from a naturally occurring group (students at Boston University, including international students). We had not anticipated the difficulty of finding people of East Asian heritage with no or little tone language exposure. Prior studies have used Japanese and Korean speakers for this category, but our results indicate that Japanese and Korean speakers had advantages that are apparently attributable to language experience, not East Asian genetics. It could be useful in future research to obtain data on sensitivity to tone in a statistical learning task from participants with fluency in Indo-European pitch-accented languages, such as Swedish and Norwegian. Future research needs to go beyond the naturalistic sample to recruit non-Asians with East Asian language experience, especially those who have become fluent or have had early exposure. Clearly, however, we are still some way from figuring out the complex interrelated effects of ethnicity and language experience.

Received 17 October 2014

NOTES

1. Here, and throughout, we use accents to denote different tones. Thus, the grave accent (e.g., *à*) indicates a low tone, the macron (e.g., *ā*) a mid tone, and the acute accent (e.g., *á*) a high tone.

2. Given the redundancy of syllabic and tonal properties, it was also possible to distinguish between words and nonwords on the basis of the ordering of tonal properties. However, as T. Wang and Saffran (2014) did not run any control conditions in which participants were only exposed to the segmental information, it is unclear if the tonal properties had any impact on learning. Instead, given that the monolingual English speakers, as a group, failed to perform above chance, the unusual tonal properties may have interfered with learning (something that T. Wang and Saffran themselves note).

3. Our study was conceived and conducted independently of T. Wang and Saffran (2014). The results of our study were first presented in March 2012 at the Annual Conference of the American Association of Applied Linguistics in Boston, Massachusetts.

REFERENCES

Alexander, J. A., Wong, P. C. M., & Bradlow, A. R. (2005). Lexical tone perception in musicians and non-musicians. *INTERSPEECH-2005*, 397–400. Retrieved from http://www.researchgate.net/publication/221484995_Lexical_tone_perception_in_musicians_and_non-musicians/file/e0b4951c9a69386504.pdf

Bent, T., Bradlow, A., & Wright, B. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 97–103.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*, 341–345.

Bowers, J. S., Mattys, S. L., & Gage, S. H. (2009). Preserved implicit knowledge of a forgotten childhood language. *Psychological Science*, *20*, 1064–1069.

https://doi.org/10.1017/S0272263114000849 Published online by Cambridge University Press

Cacioppo, J., Petty, R., & Kao, C. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*, 116–131.
Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston, MA: Houghton-Mifflin.
Chobert, J., & Besson, M. (2013). Musical expertise and second language learning. *Brain Sciences*, *3*, 923–940.
Dediu, D., & Ladd, D. R. (2007). Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, *ASPM and Microcephalin. Proceedings of the National Academy of Science*, *104*, 10944–10949.
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. G. (2006). Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period. *Journal of the Acoustical Society of America*, *119*, 719–722.
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.
Hove, M. J., Sutherland, M. E., & Krumhansl, C. L. (2010). Ethnicity effects in relative pitch learning. *Psychonomic Bulletin & Review*, *17*, 310–316.
Hyman, L. M. (2009). How (not) to do a phonological typology: The case of pitch-accent. *Language Sciences*, *31*, 213–238.
Kempe, V., & Brooks, P. J. (2008). Second language learning of complex inflectional systems. *Language Learning*, *58*, 703–746.
Kempe, V., Brooks, P. J., & Kharkhurin, A. (2010). Cognitive predictors of generalization of Russian grammatical gender categories. *Language Learning*, *60*, 127–153.
Ladd, D. R., Turnbull, R., Browne, C., Caldwell-Harris, C. L., Ganushchak, L., Swoboda, K., … Dediu, D. (2013). Patterns of individual differences in the perception of missing-fundamental tones. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 1386–1397.
Lee, Y., Vakoch, D. A., & Wurm, L. G. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, *25*, 527–542.
Liu, Y., Wang, M., Perfetti, C. A., Brubaker, B., Wu, S., & MacWhinney, B. (2011). Learning a tonal language by attending to the tone: An in vivo experiment. *Language Learning*, *61*, 1119–1141.
Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy*, *10*, 241–265.
Mattock, K., Molnar, M., Polka, L., & Durnham, D. (2007). The developmental course of lexical tone in perception in the first year of life. *Cognition*, *106*, 1367–1381.
Morett, L. M., & Chang, L.-Y. (2014). Emphasising sound and meaning: Pitch gestures enhance Mandarin lexical tone acquisition. *Language, Cognition and Neuroscience*. Advance online publication. doi:10.1080/23273798.2014.923105
Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in speech processing. *Journal of Memory and Language*, *53*, 225–237.
Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, *298*, 604–607.
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. *Journal of the Acoustical Society of America*, *130*, 461–472.
Saffran, J. R., Aslin, E., Newport, E. L. (1996). Statistical learning by 8-month old infants. *Science*, *274*, 1926–1928.
Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*, 27–52.
Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*, 606–621.
Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H. J.,... Rupp, A. (2005). Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. *Nature Neuroscience*, *8*, 1241–1247.
Schön, D., & François, C. (2011). Musical expertise and statistical learning of musical and linguistic structures. *Frontiers in Psychology*, *2*, Article 167.
So, C. K., & Best, C. T. (2011). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech*, *53*, 273–293.

https://doi.org/10.1017/S0272263114000849 Published online by Cambridge University Press

Wang, T., & Saffran, J. R. (2014). Statistical learning of a tonal language: The influence of bilingualism and previous linguistic experience. *Frontiers in Psychology*, *5*, Article 953.
Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *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.
Wayland, R. P., & Li, B. (2008). Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics*, *36*, 250–267.
Wong, P. C. M., Chandrasekaran, B., & Zheng, J. (2012). The derived allele of *ASPM* is associated with lexical tone perception. *PLoS ONE*, *7*(4), Article e34243.
Wong, P. C. M., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*, 565–585.
Yip, M. (2002). *Tone*. Cambridge, UK: Cambridge University Press.

https://doi.org/10.1017/S0272263114000849 Published online by Cambridge University Press