Musicians are better at learning non-native sound contrasts even in non-tonal languages

Amy Perfors (amy.perfors@adelaide.edu.au)
Jia Hoong Ong (jia.h.ong@student.adelaide.edu.au)
School of Psychology, University of Adelaide, Australia

Abstract

It is very difficult for adults to perceive phonetic contrasts in their non-native language. In this study we explored the effects of phonetic training for different populations of people (musicians and non-musicians) and with different kinds of phoneme contrast (timing-based, like the Hindi /g/-/k/ contrast, and pitch-based, like the Mandarin /ī/-/í/ tonal contrast). We found that musicians had superior perception for both contrasts, not just the pitch-based one. For both phonemes, training had little to no effect. We consider the implications of this for first and second language acquisition. **Keywords:** phonetic learning; music perception; language acquisition;

Introduction

Second language learning is a difficult task for a variety of reasons. Adults have difficulty with many aspects of language acquisition, including language processing (Clahsen & Felser, 2006) and certain aspects of syntax (e.g., Birdsong, 2006), but limitations in phonetic perception relative to infants and young children are especially strong and well-documented (e.g., Werker & Lalonde, 1988; Kuhl, 2004; Maye, Weiss, & Aslin, 2008). Phonological deficits can be found not just in perception, but in production and processing as well (e.g., Flege, 1995; Sebastián-Gallés & Soto-Faraco, 1999). Such deficits are sometimes thought to have cascading effects onto other aspects of language (Perani, 2005; Werker & Yeung, 2005; Perfors & Dunbar, 2010).

One striking aspect of adults' poor phonetic perception is that it is quite difficult to overcome it through training. There are various training regimes for teaching adults to learn a phonetic contrast that does not exist in their native language. Some rely on implicit learning of the phonemic categories based on distributional information (Maye & Gerken, 2001, 2002; Shea & Curtin, 2005; Hayes-Harb, 2007), while in others some form of feedback is given (e.g., Jamieson & Morosan, 1989; McCandliss, Fiez, Protopapas, Conway, & McClelland, 2002; Wang, Jongman, & Soreno, 2003; Golestani & Zatorre, 2004; Wayland & Li, 2008; Bradlow, 2008). These training regimes have rarely been compared directly, and are often used for different kinds of phonemes and with different goals. It is therefore still unclear precisely to what extent different kinds of training are effective and why.

Given the importance of phonetic perception, understanding why phonetic training works (to the extent it does) and how it can be improved (to the extent it doesn't) is a matter of some importance. One of the ways to explore this is by investigating the effectiveness of training on different kinds of phonemes as well as in different populations of people. Expanding this exploration is one of the central goals of this paper. We perform two main manipulations, comparing the effectiveness of the same implicit distributional training method on different populations (musicians and non-musicians) as well as different phonemes (tonal and timing-based). We find that musicians show improved phonetic perception on all phonemes, and that any effects of training are smaller than these population differences. The implications of these findings for language acquisition and representation more broadly are considered in the discussion.

Different phonemes, different populations

Tonal vs timing-based phonemes. A phoneme is the smallest unit in a language that forms a meaningful contrast, like the /b/ in *bat* and the /p/ in *pat*. Many phonemes are distinguished from another based on timing. For instance, one of the differences[1] between the English "g" and "k" sounds is the presence of voicing (i.e., the vibration of the vocal cords). English "g" and "k" differ in their voice onset time, or VOT, which refers to the time at which voicing begins and the vocal cords begin to vibrate. For the English "g" sound, voicing is immediate as soon as the tongue leaves the roof of the mouth; for "k", there is a time delay between the release of the stop closure and the vibration of the vocal cords.

Hindi makes a further timing-based distinction that does not exist in English. The Hindi /g/ and /k/ differ according to the presence of *pre-voicing*, which is the occurrence of voicing during the silent interval during which the vocal tract is blocked. Because this distinction does not occur in English, both sound like a "g" to a native English speaker. It is possible to train English speakers to hear this distinction, although they are far below native-speaker proficiency even after the best training. One of the simplest techniques, though not the most common for adults, is implicit distributional training (Hayes-Harb, 2007; Maye & Gerken, 2000; Maye et al., 2008). In this type of training, described in more detail later, participants hear a bimodal distribution of phonemes whose peaks are centered around the two phonemes to be learned. Distributional training has been used to teach adults to distinguish these Hindi phonemes given as little as 10 minutes of exposure (Maye & Gerken, 2000; Perfors & Dunbar, 2010).

Another kind of phoneme occurs in tonal languages like Mandarin, which uses pitch to convey meaning. In such a language, the meaning of the syllable changes when it is spoken in a different tone. For instance, in Mandarin, *ma* in the high level Tone 1 ([mā]) and the rising Tone 2 ([má]) means *mother*

[1] The other difference between these two phonemes is aspiration, which refers to the presence of a puff of air after making a sound (/g/ is not aspirated but /k/ is). We do not considering aspiration here.

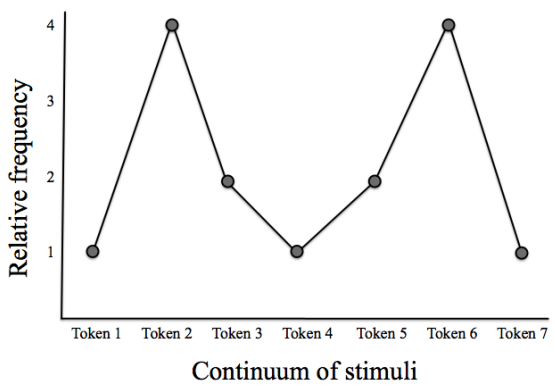

Figure 1: Distribution of stimuli used in phonetic training, defined along a continuum based on voice onset time in the HINDI condition and pitch in the MANDARIN condition. Tokens 2 and 6 occurred four times as often as tokens 1 and 7.

and *hemp*, respectively. People who are not native speakers of tonal languages process tone differently (Gandour, 1983). As with Hindi phonemes, there is evidence that non-native speakers can be trained to perceive lexical tones (Wang et al., 2003; Wayland, Herrera, & Kaan, 2010; Wong & Perrachione, 2007). However, these training programs generally take longer (days to weeks) and are more intensive than implicit distributional training, with participants being given explicit feedback and sometimes visual help (e.g., pitch graphs).

To our knowledge, no studies have explored the effectiveness of implicit distributional training on lexical tones. Our first goal in this study is therefore to compare the effects of distributional training (and baseline perception) of a Hindi timing-based contrast with a Mandarin tone-based contrast. Are both phoneme types equally easy for native English speakers to perceive? Does implicit training work better or worse with one kind of phoneme?

Musicians vs non-musicians. It is well-known that musicians consistently show a superior ability to learn lexical tones (Alexander, Wong, & Bradlow, 2005; Wong & Perrachione, 2007; Wayland et al., 2010). In many ways, this is no surprise: because both music and tonal languages involve pitch, extensive musical training (or superior auditory abilities) may result in increased sensitivity to pitch-related cues.

Much less is known about whether musicianship facilitates the learning of non-native contrasts that are not defined by pitch. While there are few studies investigating this issue, early evidence suggests that it might. For instance, Slevc and Miyake (2006) found that musical ability predicted Japanese speakers' ability to discriminate and produce the English /r/-/l/ contrast, and Sadakata, van der Zanden, and Sekiyama (2010) found that Japanese musicians were better than non-musicians at distinguishing the Dutch vowel /u/. There is also some evidence suggesting that musicians have higher brainstem plasticity not just for musical stimuli, but for speech stimuli as well (Musacchia, Sams, Skoe, & Kraus, 2007).

For all these reasons, it seems reasonable to think that people with musical training might have superior perception of timing-based phonetic contrasts as well as tones. However, to our knowledge this question has not been investigated before. Our second goal is therefore to compare the performance of musicians and non-musicians on the Hindi /g/-/k/ contrast. Are musicians better at perceiving that contrast as well as a Mandarin tonal contrast? How much does musicianship help (if it does) in either case? Are musicians more or less responsive to distributional training than non-musicians?

Goals of the study

This paper addresses two main questions. First, we are interested in comparing performance on two different kinds of phonetic contrast within people given the exact same implicit distributional training. Is one easier than the other? Does training have more of an effect for one than another? Second, we are interested in comparing different populations of adults in their ability to perceive these two kinds of contrasts: namely, musicians and non-musicians. Do musicians have an advantage in perceiving timing-based contrasts as well as pitch-based ones? Does training have more or less of an effect on them than non-musicians?

Method

There were two phases in this experiment, a training phase and a testing phase. Participants were randomly allocated to either a HINDI or MANDARIN condition. During the training, the participants were exposed to a distribution of sounds from the appropriate language for their condition. All participants participated in the same testing phase, which included two common tests of phoneme discrimination. The testing phase included stimuli from both Mandarin and Hindi, as well as a control set of English phonemes to ensure that they were paying attention. This design allows participants from each condition to serve as each other's control. For instance, the performance of participants in the HINDI condition on Mandarin stimuli reflects performance on those stimuli without having been trained on the Mandarin contrast.

Participants. 96 native English-speaking adults from the University of Adelaide and surrounding community participated in the experiment. 48 were classified as musicians according to criteria adapted from Wong and Perrachione (2007) and later slightly loosened in order to recruit enough participants. Musicianship in this study was defined as having had at least five continuous years of formal musical training, starting before the age of 15. The mean duration of musical training was 12.75 years for the musicians and 0.89 years for the non-musicians. There was no significant difference in duration of musical training between the HINDI and MANDARIN conditions ($t(46) = 0.81, p = 0.4245$, two-tailed).

Training. Participants in the HINDI condition were trained on a distribution of seven stimuli that differed according to voice onset time, while those in the MANDARIN condition were trained on a distribution of seven stimuli that differed according to pitch. The stimuli are described in detail below, but the training procedure is the same in all conditions. As

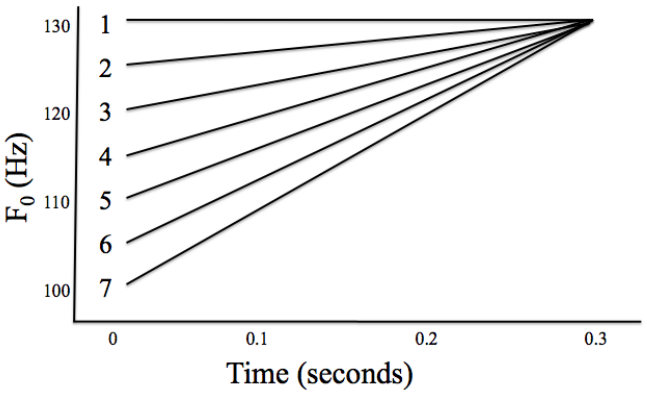

Figure 2: A seven-step continuum of the Mandarin syllable [i]. Each step differed according to its fundamental frequency (f_0) contour. (Figure adapted from Xu et al., 2006)

in Maye and Gerken (2000), we presented subjects with a bimodal distribution of these phonemes, as illustrated in Figure 1; thus, some tokens (e.g., 2 and 6) occurred four times as often as others (e.g., 1 and 7). Participants heard a total of 900 tokens presented in random order and separated by 250 ms each, for a total of approximately 10 minutes of exposure. During stimulus presentation the participants were told not to speak or read, but also that they need not consciously concentrate on the sounds. To alleviate boredom, they were allowed to doodle while listening.

Testing. After the training, the participants were presented with two standard discrimination tests, ABX and same/different (S/D); the order of the tests was randomized for each participant. In the ABX test, the participants heard three sounds separated by 1s and were asked whether the third ("X") was the same as the first ("A") or the second ("B"). In the S/D test, they heard a pair of sounds separated by 500ms and were asked whether the sounds were the same or not. Performance is given by the percentage of correct answers, and chance is 50% for both kinds of tests. Only tokens 1 and 7 were used in testing, since they most resemble the natural sounds of the language. Both tests consisted of 40 trials presented in random order, with 8 control stimuli corresponding to an English contrast, 16 stimuli corresponding to the Hindi /g/-/k/ contrast, and 16 corresponding to the Mandarin /ī/-/í/ contrast. The stimuli are described in more detail below.

Stimuli. Each of the two conditions were trained on distributions of stimuli taken from their respective languages – the HINDI condition on a contrast defined by timing, and the MANDARIN condition defined by pitch.

HINDI. The contrast used in this study was the unaspirated velar plosive voiced/voiceless contrast (/g/-/k/), which occurs in Hindi but not in English (both phonemes sound like a "g" to an English speaker). The /g/ and /k/ phonemes differ in terms of voice-onset time (VOT), such that /g/ contains a prevoicing component while /k/ does not. It is therefore possible to gradually convert /g/ tokens into /k/ by successively removing parts of the pre-voicing component. Our training stimuli consisted of the Hindi syllable pairs [gi]-[ki], constructed by recording a male native Hindi speaker saying [gi] and systematically removing the pre-voicing component using Praat phonetics software. This yields a continuum of seven stimuli from [gi] to [ki], separated by an average of 19ms in VOT from each other, and identical except for the pre-voicing.

Half of the 16 test trials used the [gi]-[ki] stimuli, with order of presentation of each and the side of the correct response counterbalanced. The other half of the test trials (also fully counterbalanced) consisted of the same contrast spoken in a different vowel context ([ga]-[ka]) and recorded by a female native Hindi speaker. The continuum of [ga] to [ka] was constructed in the same way as [gi] to [ki], although only tokens 1 and 7 were used during the test trials. For space reasons, we will report on overall performance among all of the test trials rather than on the two kinds of test trials individually.

MANDARIN. Participants were trained on a continuum bridging two tones, one of which is high level (Tone 1) and one of which is rising (Tone 2), as obtained from Xu, Gandour, and Francis (2006) and illustrated in Figure 2. The training stimuli consisted of a continuum between a vowel in Tone 1 ([ī]) and the same vowel in Tone 2 ([í]). A male native Mandarin speaker produced the syllable [ī], and its fundamental frequency (f_0) contour was systematically altered to synthesize tokens on the continuum to Tone 2. This resulted in a seven-equal step continuum, as shown in Figure 2. All tokens have the same offset frequency (130Hz) and were normalized for duration and amplitude. As documented in Xu et al. (2006), three native Mandarin speakers judged tokens 1 and 7 to be good exemplars of the [ī] and [í] syllables.

Analogously to the Hindi stimuli, half of the 16 Mandarin test trials used the [ī]-[í] stimuli, with the order and side fully counterbalanced. The other half of the test trials consisted of the same tonal contrast spoken by a female Mandarin speaker with a different vowel ([ā] to [á]), constructed by the same method as the [ī]-[í] stimuli. As with the Hindi stimuli, we will report on overall performance among all of the test trials rather than on the two kinds of test trials individually.

CONTROL. In order to make sure that participants were attending to and understood the task, we included 8 trials of control stimuli during each of the two tests. These corresponded to a phonemic contrast they could already recognize: the dental plosive aspirated/unaspirated voiced/voiceless contrast (/d/-/t^h/, which sound like "d" and "t" respectively to a native English speaker). Because the /d/-/t^h/ contrast also exists in Hindi, the phonemes were recorded by the same male Hindi speaker as before.

Results

Our study used two phoneme discrimination tests, the ABX and the S/D. Performance on these two tests was comparable: there was no significant difference in overall percent correct between discrimination tests (paired-sample $t(383) = 1.33, p = 0.184$), with a mean performance of 70.5% (SD=20.4) in the ABX tests and a mean performance of 69.2% (SD=18.8) in the S/D tests. Moreover, the scores on

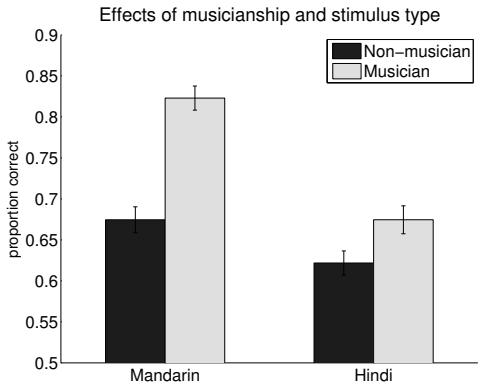

Figure 3: Overall accuracy by musicianship and stimulus type. Musicians showed superior performance to non-musicians, and performance was also higher on Mandarin than Hindi stimuli. Moreover, there was an interaction indicating that being a musician helped relatively more for Mandarin than Hindi stimuli.

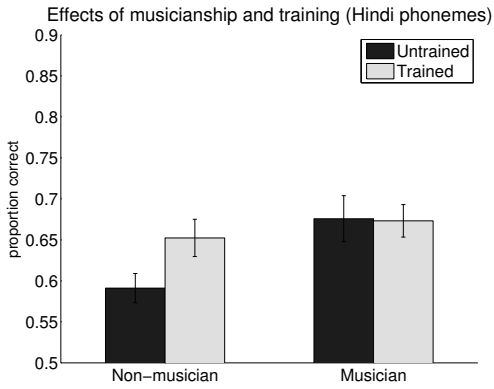

Figure 4: Overall accuracy on the Hindi phonemes by musicianship and training. Musicians did significantly better than non-musicians. Training did not significantly improve performance, although the trend among non-musicians approached significance.

the two tests were correlated ($r = 0.50, p < 0.0001$). In all of the subsequent analyses we therefore collapse performance on the two tests into one overall accuracy score.

Our first question is whether discrimination performance is different on the Hindi and the Mandarin contrasts: is one easier than the other? Is there a differential effect of musicianship on each? We address this by considering performance on the stimuli for each language, collapsing (for now) any effects of training. As Figure 3 shows, musicians performed better than non-musicians for both types of contrast (two-way ANOVA, $F(1,380) = 41.5, p < 0.0001$) and performance was higher on the Mandarin stimuli ($F(1,380) = 41.5, p < 0.0001$). Moreover, there was an interaction, indicating that being a musician helped more for Mandarin stimuli than for Hindi stimuli ($F(1,380) = 9.39, p = 0.002$).

How did the effects of musicianship play out within each stimulus type, and how did that interact with training? Were musicians or non-musicians helped more by training? Were the effects different depending upon the nature of the contrast in question? To address these issues we evaluate performance on the Hindi and Mandarin stimuli separately.

Hindi stimuli. Figure 4 shows overall accuracy on the Hindi test stimuli by musicianship and training. Participants were considered to have been trained on the stimuli if they were in the HINDI condition and untrained if they were in the MANDARIN condition. Musicians performed significantly better than non-musicians (two-way ANOVA, $F(1,188) = 5.53, p = 0.019$), but there was no significant effect of training ($F(1,188) = 1.71, p = 0.193$) and no interaction ($F(1,188) = 0.049, p = 0.156$).

Mandarin stimuli. Figure 5 shows overall accuracy on the Mandarin test stimuli by musicianship and training (where people in the HINDI condition were considered to be untrained on the Mandarin stimuli). Here too, musicians performed significantly better than non-musicians (two-way ANOVA, $F(1,188) = 46.8, p < 0.0001$). As before, there was no significant effect of training ($F(1,188) = 0.001, p = 0.810$) and no interaction ($F(1,188) = 0.71, p = 0.402$).

Discussion

Our overall findings show a strong effect of musicianship on phonetic perception: musicians had superior perception for both timing-based and pitch-based non-native contrasts (although the effect was much stronger for pitch-based contrasts). Interestingly, there was no effect of distributional training. Here we consider some of the implications of these findings for first and second language acquisition in general.

Why did musicians perform better?

These results are consistent with the extensive literature documenting the fact that musicians have superior performance in linguistic tasks involving lexical tones (Alexander et al., 2005; Wong & Perrachione, 2007; Wayland et al., 2010). However, there is relatively little prior work showing that musicians have an advantage for non-tone-based phonemes, and none that we know of that investigates phonemes defined by differences in voice onset time (Slevc & Miyake, 2006; Sadakata et al., 2010). This suggests that whatever advantage musicians enjoy is not limited to differences in pitch perception, even though the timing-based advantages are smaller. Indeed, we even found a slight difference in performance between musicians and non-musicians on the control stimuli; although both groups performed extremely well (98.7% accuracy for the musicians, 95.2% accuracy for the non-musicians), the difference between the groups was significant ($t(94) = 2.94, p = 0.004$). This is somewhat surprising, but it is true that even the control sounds were potentially confusable than more distinct phonemes would have been, and the difference is small in magnitude.

What is the root of the musician advantage? Consistent with their slightly better performance even on the control trials, one possibility is that musicians simply have a "better ear" in general – that is, they have superior auditory processing abilities overall. While this possibility is consistent with existing research (Schön, Magne, & Besson, 2004; Wong, Skoe, Russo, Dees, & Kraus, 2007; Musacchia et al., 2007), it does not really answer the question: *why* do they have superior abilities? Does musical training itself improve such

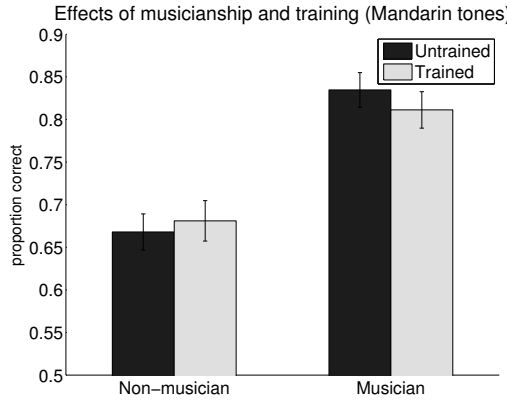

Figure 5: Overall accuracy on the Mandarin phonemes by musicianship and training. Musicians did significantly better than non-musicians, and training had no effect on performance.

abilities, or is it simply that people with better auditory processing become musicians in the first place?

This is a hard question to firmly disentangle, but our data provides one way to address it. We find that although the total duration of musical training is correlated with overall accuracy ($r = 0.306, p = 0.002$), this effect is carried by the presence of non-musicians in the sample; there is no effect of duration of musical experience among the musicians only ($r = 0.012, p = 0.932$). This occurs despite the fact that our sample of musicians was fairly diverse, ranging from people with 5 years to 45 years of training (M = 12.75, SD=7.7). It implies that perhaps at least part of the difference between musical and non-musical populations may be due to non-training-related differences in auditory perception. That said, since our participants played a wide range of instruments and were involved with music at different intensities – and any effects of duration are confounded with age – this is at best suggestive and should be interpreted with caution.

Could the better performance by the musicians be due to a motivational difference? While this might explain why the musicians performed better even on the control trials, it seems unlikely. The musicians were not told explicitly that they were being compared to non-musicians; they were only told that we were interested in how musicians learn language in general. It is unclear why they would be more motivated in such a scenario. Furthermore, our findings are consistent with the large amount of previous work showing superior phonetic processing in musicians (Alexander et al., 2005; Slevc & Miyake, 2006; Sadakata et al., 2010; Wayland et al., 2010). That said, in follow-up work we plan to investigate this possibility by recruiting visual artists, who would have the same motivational advantages (if any) that come from being recruited as a member of a special group, but who we would not expect to have superior auditory perceptual abilities.

Why did training have no effect?

In addition to the musician advantage, our other main finding was that distributional training had no effect on the discrimination of phonemes in our experiment.[2] Although a few previous experiments (Maye & Gerken, 2000; Hayes-Harb, 2007; Perfors & Dunbar, 2010) have found effects of distributional training for Hindi contrasts among non-musicians, these effects were small. Indeed, most training regimes for adults last far longer than 10 minutes and give reinforcement of some type, precisely because it is far more effective to do so(Jamieson & Morosan, 1989; McCandliss et al., 2002; Golestani & Zatorre, 2004; Bradlow, 2008).

The lack of training effect among musicians and for the Mandarin stimuli may have occurred for similar reasons. However, it also may be that implicit distributional training is not an effective means for teaching adults to discriminate pitch-based contrasts like the Mandarin /ī/-/í/ distinction. This seems more plausible given the fact that even within the non-musicians – for whom there was non-significant trend of training for the Hindi contrast – there was no effect of training on the Mandarin contrast. But why would distributional training be more effective for one kind of contrast than another? We know that it is *possible* to train non-native speakers to perceive this kind of tonal contrast (Wang et al., 2003; Wayland et al., 2010; Wong & Perrachione, 2007). However, the training programs that have been successful have been far more intensive and explicit than ours was. We can only speculate, but perhaps this sort of instruction is necessary for people to understand the pitch distinctions they should be listening for. In contrast, distributional training may naturally focus people's attention on timing by playing one phoneme after another in rapid succession. That said, since distributional training had no statistically significant effect in any case, the simplest conclusion is that it was insufficient regardless of the nature of the contrast.

A related possibility is that our participants were already performing near their ceiling – that is, near the peak of what would be possible for them without decades of experience distinguishing the sounds in question. Perhaps the benefits of training for non-musicians in Hindi found in previous studies and implied here come from making people aware of the more obvious cues that can be used to distinguish the sounds. Due to their superior auditory skills, musicians may already be aware of those cues; and pitch differences are blatant enough that even non-musicians can hear some of the differences between /ī/ and /í/. Of course, the difference between musicians and non-musicians, and between Hindi and Mandarin phonemes, also implies that if there is a ceiling effect, there are probably multiple different ceilings rather than only one.

Is there a ceiling effect? Given the widely-documented difficulty of training adults to recognize non-native contrasts,

[2] We did find that if we performed a post-hoc analysis on only non-musicians in the Hindi contrast, which is the only condition directly corresponding to the previous literature (Maye & Gerken, 2000; Hayes-Harb, 2007; Perfors & Dunbar, 2010), the training effect was significant ($t(94) = 2.12, p = 0.037$). This suggests that had that been the only condition studied – analogous to the previous literature – it would have reached significance. However, since this analysis in our study followed an omnibus ANOVA with no main effect or interaction, it is not statistically appropriate to apply here.

this is certainly possible. After all, native speakers have decades of experience distinguishing between those sounds, and it would be quite surprising if these vast differences in exposure could be eliminated with a small amount of training, even among especially capable subjects like musicians. The notion of a ceiling effect for training is also consistent with the Native Language Neural Commitment hypothesis, which suggests that early experience results in changes in the brain that encode the phonetic contrasts of one's native language (Kuhl, 2004). Because of these neural changes, learning non-native contrasts as an adult is therefore extremely difficult. This would also explain why training made no difference among any of the musicians, who may have been already performing near the limit possible for brains that grew up dedicated to hearing other kinds of contrasts. If there is a ceiling, it has unfortunate implications for second language acquisition. Perhaps it is intrinsically limited by poor phonetic perception abilities, at least without years and years of exposure to the new language. That said, we must remind ourselves that the training approach used in this study was quite simple and short compared to other, more intensive methods, which might very well have more of an effect.

In many ways, this study raises more questions than it answers. Why did distributional training have no effect, particularly for musicians and Mandarin contrasts? What is the root of the musician advantage, and why does it extend to include timing-based contrasts as well as pitch-based ones? These questions are still open, but this work is an important step toward understanding the roots of phonetic perception and its relationship to both first and second language learning.

Acknowledgments

Thank you to Natalie May, Angela Vause, and Daniel Carabellese for recruiting participants and running the experiments, and to Dan Navarro and the CLCL lab for useful discussions. We especially thank Dr. Jackson Gandour for generously sharing his Mandarin stimuli. AP was supported by ARC grant DE120102378.

References

Alexander, J., Wong, P., & Bradlow, A. (2005). Lexical tone perception in musicians and nonmusicians. In *9th European Conference on Speech Comm. and Tech.* Lisbon.

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning*, *56*(1), 9–49.

Bradlow, A. (2008). Training non-native language sound patterns. In J. Hansen Edwards & M. Zampini (Eds.), *Phonology and second language acquisition* (p. 287-308). Benjamins.

Clahsen, H., & Felser, C. (2006). How native-like is non-native language processing? *Trends in Cognitive Sciences*, *10*(12), 564–570.

Flege, J. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 229–273). Timonium, MD: York Press.

Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, *11*, 149-175.

Golestani, N., & Zatorre, R. (2004). Learning new sounds of speech: Reallocation of neural substrates. *NeuroImage*, *21*, 494-506.

Hayes-Harb, R. (2007). Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research*, *23*(1), 65–94.

Jamieson, D., & Morosan, D. (1989). Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques. *Canadian Journal of Psychology*, *43*(1), 88–96.

Kuhl, P. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*, 831–843.

Maye, J., & Gerken, L. (2000). Learning phonemes without minimal pairs. In *24th Annual Meeting of the Boston Uuniversity Conference on Language Development.*

Maye, J., & Gerken, L. (2001). Learning phonemes: How far can the input take us? In *25th Annual Meeting of the Boston Uuniversity of the Conference on Language Development.*

Maye, J., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*, B101-B111.

Maye, J., Weiss, D., & Aslin, R. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, *11*(1), 122–134.

McCandliss, B., Fiez, J., Protopapas, A., Conway, M., & McClelland, J. (2002). Success and failure in teaching the [r]-[l] contrast to japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, & Behavioral Neuroscience*, *2*(2), 89–108.

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences*, *104*(40), 15894-15898.

Perani, D. (2005). The neural basis of language talent in bilinguals. *Trends in Cognitive Sciences*, *9*(5), 211–213.

Perfors, A., & Dunbar, D. (2010). Phonetic training makes word learning easier. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society.*

Sadakata, M., van der Zanden, L., & Sekiyama, K. (2010). Influence of musical training on perception of l2 speech. In *11th Annual Conference of the International Speech Communication Association.* Chiba.

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

Sebastián-Gallés, N., & Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition*, *72*, 111–123.

Shea, C., & Curtin, S. (2005). Learning allophones from the input. In *29th Annual Meeting of the Boston Uuniversity of the Conference on Language Development.*

Slevc, L., & Miyake, A. (2006). Individual differences in second language proficiency: Does musical ability matter? *Psychological Science*, *17*(8), 675-681.

Wang, Y., Jongman, A., & Soreno, J. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America*, *113*(2), 1033-1043.

Wayland, R., Herrera, E., & Kaan, E. (2010). Effects of musical experience and training on pitch contour perception. *Journal of Phonetics*, *36*, 250-267.

Wayland, R., & Li, B. (2008). Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics*, *36*, 250-267.

Werker, J., & Lalonde, C. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology*, *24*(5), 672–683.

Werker, J., & Yeung, H. (2005). Infant speech perception bootstraps word learning. *Trends in Cognitive Sciences*, *9*(11), 519–527.

Wong, P., & Perrachione, T. (2007). Learning pitch patterns in lexical identification by native english-speaking adults. *Applied Psycholinguistics*, 565-585.

Wong, P., Skoe, E., Russo, N., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, *10*, 420-422.

Xu, Y., Gandour, J., & Francis, A. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *Journal of Acoustical Society of America*, *120*(2), 1063-1074.