Check for updates

Concurrent Sound Segregation Is Enhanced in Musicians

Benjamin Rich Zendel and Claude Alain

Downloaded from http://direct.mit.edu/jocn/article-pdf/21/8/1488/1937086/jocn.2009.21140.pdf by guest on 18 May 2021

Abstract

■ The ability to segregate simultaneously occurring sounds is fundamental to auditory perception. Many studies have shown that musicians have enhanced auditory perceptual abilities; however, the impact of musical expertise on segregating concurrently occurring sounds is unknown. Therefore, we examined whether long-term musical training can improve listeners' ability to segregate sounds that occur simultaneously. Participants were presented with complex sounds that had either all harmonics in tune or the second harmonic mistuned by 1%, 2%, 4%, 8%, or 16% of its original value. The likelihood of hearing two sounds simultaneously increased with mistuning, and this effect was greater in musicians than nonmusicians. The segregation of the mistuned harmonic from the harmonic series was paralleled by an object-related negativity that was larger and peaked earlier in musicians. It also coincided with a late positive wave referred to as the P400 whose amplitude was larger in musicians than in nonmusicians. The behavioral and electrophysiological effects of musical expertise were specific to processing the mistuned harmonic as the N1, the N1c, and the P2 waves elicited by the tuned stimuli were comparable in both musicians and nonmusicians. These results demonstrate that listeners' ability to segregate concurrent sounds based on harmonicity is modulated by experience and provides a basis for further studies assessing the potential rehabilitative effects of musical training on solving complex scene analysis problems illustrated by the cocktail party example. ■

INTRODUCTION

Musical performance requires rapid, accurate, and consistent perceptual organization of the auditory environment. Specifically, this requires the organization of acoustic components that occur simultaneously (i.e., concurrent sound organization) as well as the organization of successive sounds that takes place over several seconds (i.e., sequential organization). Broadly, this organization of the auditory world is known as ''auditory scene analysis,'' which is important because natural auditory environments often contain multiple sound sources that occur simultaneously (Bregman, 1990). The present study focused on the impact of musical expertise on listeners' ability to perceptually organize sounds that occur concurrently.

A powerful way to organize the incoming acoustic waveform is based on the harmonic relations between components of a single physical sound source. If a tonal component is not harmonically related to the sound's fundamental frequency (f_0), it can be heard as a simultaneous but separate entity, especially if it is a lower rather than a higher harmonic and if the amount of mistuning is greater than 4% of its original value (Alain, 2007; Moore, Glasberg, & Peters, 1986). The mechanisms underlying the perception of the mistuned harmonic as a separate sound are not well understood but likely involve neurons that are sensitive to frequency periodicity. Neurophysiological studies indicate that violations of harmonicity (i.e., a mistuned harmonic) are registered at various stages along the ascending auditory pathways including the auditory nerve (Sinex, Guzik, & Sabes, 2003), the cochlear nucleus (Sinex, 2008), the inferior colliculus (Sinex, Sabes, & Li, 2002), and the primary auditory cortex (Fishman et al., 2001). These early and automatic representations of frequency suggest that violations of harmonicity are encoded as primitive cues to parsing the auditory scene.

In humans, the neural correlates of concurrent sound processing have been investigated using scalp recorded ERPs. When ERPs elicited by a complex sound are compared with those elicited by the same complex sound with a mistuned tonal component (especially above 8%), an increased negativity is observed, which peaks around 140 msec poststimulus onset (see Alain, 2007). This object-related negativity (ORN) is best illustrated by subtracting ERPs to tuned stimuli from those elicited by the mistuned stimuli. The difference wave reveals a negative deflection at fronto-central sites that reverses in polarity at electrodes placed near the mastoids and the cerebellar areas.

The segregation of concurrent sounds based on harmonicity, as indexed by ORN generation, is little affected by attentional demands, as ORN can be observed in situations where participants are attending to other tasks including a contralateral auditory task (Alain & Izenberg, 2003), reading a book (Alain, Arnott, & Picton, 2001), or

University of Toronto

© 2008 Massachusetts Institute of Technology

watching a silent movie (Alain, Schuler, & McDonald, 2002). These findings provide strong support for the proposal that the organization of simultaneous auditory objects is not under volitional control. However, when participants were asked to make a perceptual judgment about the incoming complex sounds (i.e., if he or she heard one sound or two simultaneous sounds), the likelihood of reporting two concurrent sounds was correlated with ORN amplitude (see Alain, 2007). In addition, when subjects reported hearing two simultaneous sounds, a later positive difference (tuned minus mistuned stimuli) wave peaking at about 400 msec after sound onset (P400) emerged (Alain et al., 2001). Like the ORN, the amplitude of the P400 correlated with perceptual judgment, being larger when participants perceived the mistuned harmonic as a separate tone (Alain et al., 2001). These findings suggest that the P400 reflects a conscious evaluation and decision-making process regarding the number of auditory objects present, whereas the ORN reflects low-level primitive perceptual organization (see Alain, 2007).

One important issue that remains unanswered and deserves further empirical work is whether the organization of simultaneous acoustic components can be enhanced by experience. It is well accepted that auditory scene analysis engages learned schema-driven processes that reflect listeners' intention, experience, and knowledge of the auditory environment (Bregman, 1990). For instance, psychophysical studies have shown that presenting an auditory cue with an identical frequency to an auditory target improved detection of the target when embedded in noise (Hafter, Schlauch, & Tang, 1993; Schlauch & Hafter, 1991). Similarly, familiarity with a melody facilitates detection when interweaved with distracter sounds (Bey & McAdams, 2002; Dowling, 1973). Hence, schema-driven processes provide a way to resolve perceptual ambiguity in complex listening situations when the signal to noise ratio is poor. In more recent studies, short-term training (over the course of an hour or a few days) has been shown to improve listeners' ability to segregate and to identify two synthetic vowels presented simultaneously in young (Alain, Snyder, He, & Reinke, 2007; Reinke, He, Wang, & Alain, 2003) as well as in older adults (Alain & Snyder, 2008), suggesting that learning and intention can enhance sound segregation and identification. However, it is unclear from these studies whether improvement in identifying concurrent vowels occurred because of a greater reliance on schema-driven processes or whether the improvement also reflects learning-related changes in primitive auditory processes.

Studies measuring scalp-recorded ERPs suggest that musical expertise may be associated with neuroplastic changes in early sensory processes. For instance, the amplitude of the N1[1] (Pantev, Roberts, Schultz, Engelien, & Ross, 2001; Pantev et al., 1998), N1c (Shahin, Bosnyak, Trainor, & Roberts, 2003), and P2 (Shahin, Roberts, Pantev, Trainor, & Ross, 2005; Shahin et al., 2003) waves, evoked by transient tones with musical timbres, are larger in musicians compared with nonmusicians. The N1 is further enhanced in musicians when the evoking stimulus is similar in timbre to the instrument on which they were trained, with violin tones evoking a larger response in violinists and trumpet tones evoking a larger response in trumpeters (Pantev et al., 2001). Similarly, increasing the spectral complexity of a sound so that it approached the sound of a real piano yielded a larger P2 wave in musicians compared with nonmusicians (Shahin et al., 2005). More importantly, these enhancements are smaller or nonexistent when presented with pure tones, suggesting that the observed changes in sensory-evoked responses in musicians are specific to musical stimuli (Shahin et al., 2005; Pantev et al., 1998). In addition to the cortical change related to processing sounds with musical timbres, evidence suggests that the encoding of frequency at the subcortical level (i.e., the brain stem) is also enhanced in musicians, which suggests that low-level auditory processing may be modulated by experience (Wong, Skoe, Russo, Dees, & Kraus, 2007).

The current study investigated whether long-term musical training influenced the segregation of concurrently occurring sounds. The nature of music performance involves the processing of multiple sounds occurring simultaneously, which leads us to believe that expert musicians should demonstrate enhanced concurrent sound segregation paralleled by modulations to the associated neural correlates. By using nonmusical stimuli, we assessed whether general (not specific to music) processes were influenced by long-term musical training. To test this hypothesis, we presented participants with complex sounds similar to those of Alain et al. (2001), and they indicated whether the incoming harmonic series fused into a single auditory object or whether it segregated into two distinct sounds, that is, a buzz plus another sound with a pure tone quality. In addition, the same stimuli were presented without requiring a response to examine whether electrophysiological differences related to musical expertise were response dependent. It was expected that the perception of concurrent auditory objects will increase as a function of mistuning and that the perception of concurrent sounds will be paralleled by ORN and P400 waves, as was found in previous studies (e.g., Alain & Izenberg, 2003; Alain et al., 2001, 2002). In addition, it was hypothesized that musicians will be more likely to report hearing the mistuned harmonic as a separate sound and that these behavioral changes will be accompanied by changes to the ORN and the P400 waves.

METHODS

Participants

Twenty-eight participants were recruited for the study: 14 expert musicians (M = 28.2 years, SD = 3.2, 8 women)

Downloaded from http://direct.mit.edu/jocn/article-pdf/21/8/1488/1937686/jocn.2009.21140.pdf by guest on 18 May 2021

and 14 nonmusicians (M = 32.9 year, SD = 9.9, 7 women). Expert musicians were defined as having advanced musical training (i.e., undergraduate or graduate degree in music, conservatory Grade 8 or equivalent) and continued to practice on a regular basis. Nonmusicians had no more then 1 year of formal or self-directed music lessons and did not play any musical instruments. All participants were screened for hearing loss and neurological and psychiatric illness. In addition, all participants had pure tone thresholds below 30 dB hearing level (HL) for frequencies ranging from 250 to 8000 Hz.

Stimuli

Stimuli consisted of six complex sounds each comprising six harmonically related tonal elements. The fundamental frequency was 220 Hz. Each component (220, 440, 660, 880, 1100, and 1320 Hz) was a pure tone sine wave generated with Sig-Gen software (Tucker-Davis Technology, Alachua, FL) and had durations of 150 msec with 10 msec rise/fall times. The pure tone components were combined into a harmonic complex using Cubase SX (Steinberg, V.3.0, Las Vegas, NV). The third component (second harmonic) of the series (660 Hz) was either tuned or mistuned by 1%, 2%, 4%, 8%, or 16%, corresponding to 666.6, 673.2, 686.4, 712.8, and 765.6 Hz, respectively. All stimuli were presented binaurally at 80 dB sound pressure level (SPL) through ER 3A insert earphones (Etymotic Research, Elk Grove).

Procedure

Stimuli were presented in two listening conditions, active and passive. A total of 720 stimulus iterations (120 exemplars of each stimulus type) were presented in each condition. During the passive condition, participants were instructed to relax and not to pay attention to the sounds being presented. The passive condition was spread across two blocks of 360 randomly ordered stimulus presentations with interstimulus intervals (ISIs) that varied randomly between 1200 and 2000 msec. The active condition was spread across four blocks of 180 stimulus presentations in random order with an ISI that varied randomly between 2000 and 3000 msec. After each trial, participants indicated whether they heard one complex sound (i.e., a buzz) or whether they heard two sounds (i.e., a buzz plus another sound with a pure tone quality) by pressing a button on a response box. The longer ISI in the active condition allowed time for a response. All participants first completed a passive block, then four active blocks, and finally a second passive block.

Recording of Electrical Brain Activity

Neuroelectric brain activity was digitized continuously from 64 scalp locations with a band-pass filter of 0.05–100 Hz and a sampling rate of 500 Hz per channel using SynAmps2 amplifiers (Compumedics Neuroscan, El Paso, TX) and stored for analysis. Electrodes on the outer canthi and at the superior and inferior orbit monitored ocular activity. During recording, all electrodes were referenced to electrode Cz; however, for data analysis, we re-referenced all electrodes to an average reference.

All averages were computed using BESA software (version 5.1.6). The analysis epoch included 100 msec of prestimulus activity and 1000 msec of poststimulus activity. Trials containing excessive noise ($\pm$125 μV) at electrodes not adjacent to the eyes (i.e., IO1, IO2, LO1, LO2, FP1, FP2, FP9, FP10) were rejected before averaging. ERPs were then averaged separately for each condition, stimulus type, and electrode site.

For each participant, a set of ocular movements was obtained before and after the experiment (Picton et al., 2000). From this set, averaged eye movements were calculated for both lateral and vertical eye movements as well as for eye blinks. A PCA of these averaged recordings provided a set of components that best explained the eye movements. The scalp projections of these components were then subtracted from the experimental ERPs to minimize ocular contamination such as blinks, saccades, and lateral eye movements for each individual average. ERPs were then digitally low-pass filtered to attenuate frequencies above 30 Hz.

All data were analyzed using a mixed design repeated measures ANOVA with musical training (musician and nonmusician) as a between-subjects factor and mistuning of the second harmonic (tuned, 1%, 2%, 4%, 8%, and 16%) as a within-subjects factor. For ERP data, condition (active and passive) and various electrode montages were included as within-subjects factors. The first analysis examined the effect of musical expertise on the peak amplitude and the latency of the N1, N1c, P2, and late positive complex (LPC). The N1 wave was defined as the largest negative deflection between 85 and 120 msec and was quantified at fronto-central scalp sites (Fz, F1, F2, FCz, FC1, FC2, Cz, C1, and C2). The N1c was defined as the maximum negative deflection between 110 and 210 msec at the left and right (T7/T8) temporal electrodes. The P2 peak was measured during the 130- and the 230-msec interval at fronto-central scalp sites (Fz, F1, F2, FCz, FC1, FC2, Cz, C1, and C2). Lastly, the LPC was quantified between 300 and 700 msec at parietal and parieto-occipital sites (Pz, P1, P2, POz, PO3, and PO4).

The second and the third analyses focus on the ORN and the P400 components, respectively. The effect of musical expertise on the ORN was quantified by comparing the mean amplitude during the 100- to 180-msec interval following stimulus onset with ANOVA, using musical expertise, listening condition, and mistuning level as factors. Two analyses were conducted over two different brain regions: The first was quantified over nine fronto-central electrodes (Fz, F1, F2, FCz, FC1, FC2, Cz, C1, and C2), and the second was quantified over four mastoid/cerebellar electrodes (M1, M2, CB1, and CB2).

Downloaded from http://mitpress.mit.edu/jocn/article-pdf/21/8/1483/1937686/jocn.2009.21140.pdf by guest on 18 May 2021

These electrodes were chosen because the peak activation of the ORN and its inversion were observed at these points. Moreover, the measurements over the left and right mastoids and the cerebellar electrodes allow us to test for potential hemispheric differences in processing the mistuned harmonic. For the P400, the effect of musical expertise was quantified for the mean amplitude during the 300- to 400-msec interval with ANOVA, using musical expertise and mistuning level as factors (condition was excluded for reasons explained below). As with the ORN, two analyses were conducted over two different brain regions. The first was quantified over a widened fronto-central scalp region to account for the right asymmetry of the P400 (Fz, F1, F2, FCz, FC1, FC2, Cz, C1, C2, C3, and C4), and the second was quantified over the left and the right mastoid/cerebellar sites (CB1, CB2, M1, and M2). In addition, the rate of change in amplitude during both of these time windows (100–180 and 300–400 msec) as a function of mistuning and musical expertise was also examined by orthogonal polynomial decomposition with a focus on the linear and quadratic trends.

Preliminary analyses indicated that the ORN recorded during the first and the second passive listening blocks were comparable. Thus, the ERPs recorded during these two blocks of trials were averaged together, and subsequent analyses were performed on the ERPs averaged across block. For the P400 wave, the effects of musical expertise and mistuning were limited to ERPs recorded during the active listening condition because there was no reliable P400 wave during the passive listening (differences between Blocks 1 and 2 were also examined, and no difference was found). Thus, all analyses on the P400 were done only during active listening.

RESULTS

Behavioral Data

Figure 1 shows the proportion of trials where participants reported hearing two concurrent sounds as a function of mistuning. The ANOVA yielded a main effect of mistuning, $F(5,130) = 133.7$, $p < .001$, and a significant interaction between expertise and mistuning, $F(5,130) = 3.68$, $p < .01$. Post hoc comparisons revealed that musicians were more likely than nonmusicians to report hearing two simultaneous sounds when the second harmonic was mistuned by 4%, 8%, and 16% ($p < .05$ in all cases). There was no difference in perceptual judgment between musicians and nonmusicians when the second harmonic was either tuned or mistuned by 1% ($p > .1$), but there was a trend toward a difference at 2% ($p = .09$).

Electrophysiological Data

Figure 2A and B show the group mean ERPs averaged across stimulus type during active and passive listening, respectively. The ERPs comprised N1 and P2 waves that were largest over the fronto-central scalp sites and peaked at about 100 and 180 msec after sound onset, respectively. During active listening, the N1–P2 complex was followed by a sustained potential that was positive and maximal over the parietal regions, referred to as an LPC. First, analyses of N1, N1c, and P2 peaks were done only on tuned stimuli to examine whether musical expertise modulates the processing of complex sounds irrespective of mistuning. The main effect of musical expertise on the N1, N1c, and P2 amplitude was not significant nor was the interaction between musical expertise and listening condition ($p > .2$ in all cases). The N1 and N1c were both larger in active listening, $F(1,26) = 30.9$ and 14.0, $p < .01$; however, the P2 was not affected by listening condition ($p > .2$).

In subsequent analyses, mistuning was included as an additional factor. As expected, the N1 and the N1c waves were larger during active than passive listening, $F(1,26) = 41.93$ and 12.08, $p < .01$, and the P2 wave was not affected by listening conditions ($p > .2$). The main effect of musical expertise and the interaction between expertise and listening condition were not significant for N1, N1c, or P2 ($p > .1$); however, the effect of mistuning interacted with musical expertise for the N1 and P2, $F(5,130) =$

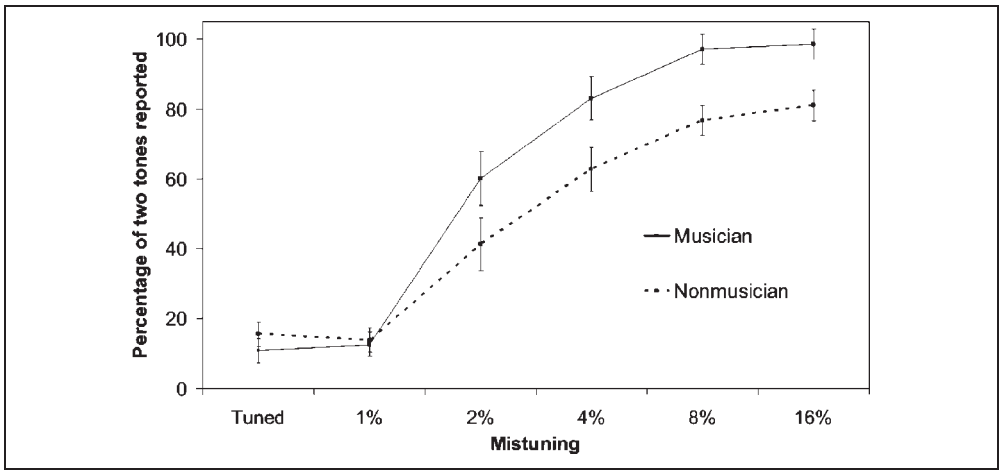

Figure 1. Percentage of stimuli perceived as two tones as a function of mistuning of the second harmonic (error bars = 1 *SE*).

Downloaded from http://direct.mit.edu/jocn/article-pdf/21/8/1488/1837186/jocn.2009.21140.pdf by guest on 18 May 2021

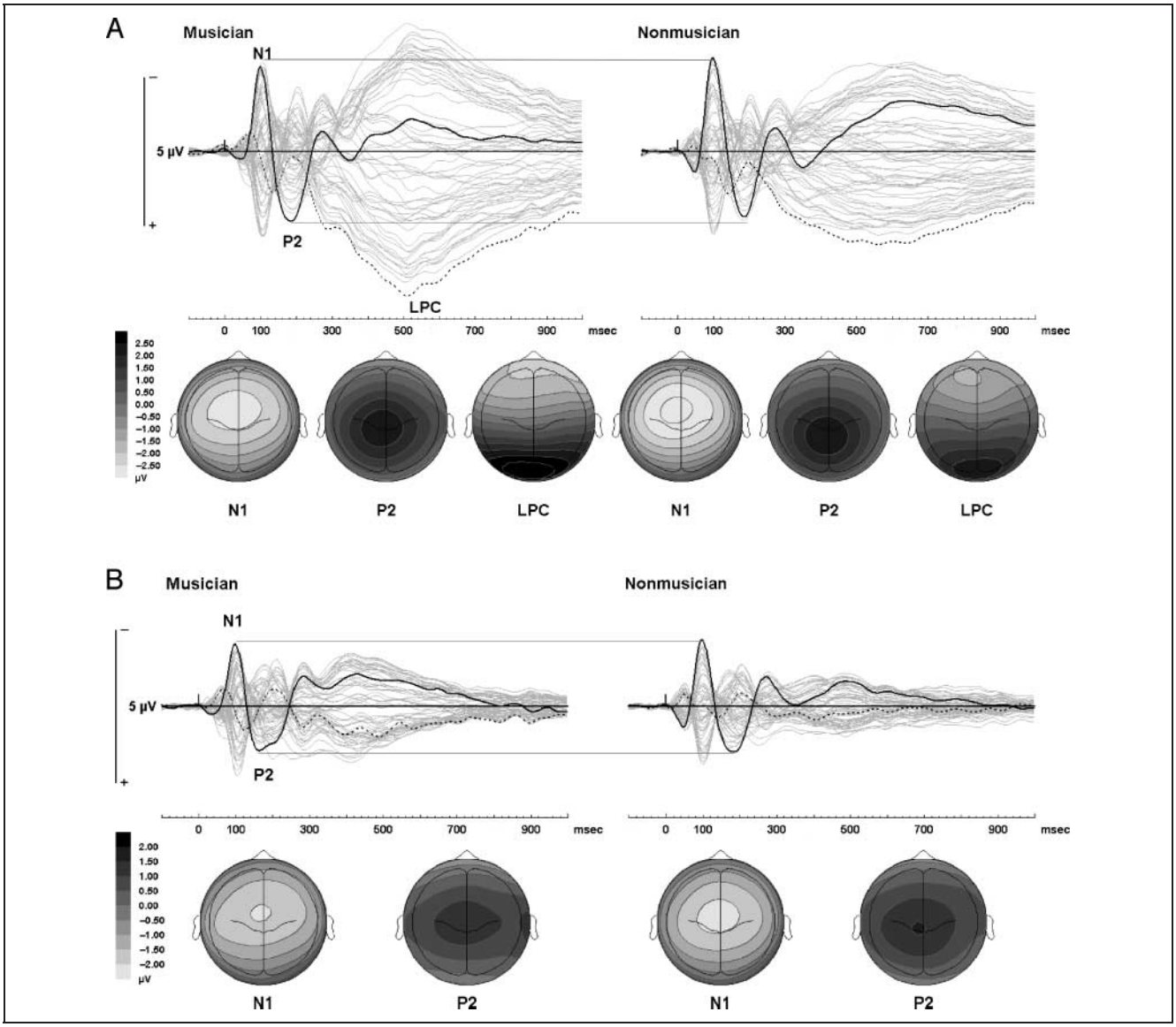

Figure 2. (A) Active listening: Sensory-evoked responses averaged across all mistuning conditions in active trials separated by group. The topographic maps for each peak show activity at the following latencies: N1, 100 msec; P2, 180 msec; and LPC, 500 msec. Electrode Cz is a solid black line, POz is a dotted line, and all other electrodes are gray. Horizontal gray lines show that the amplitude of N1 and P2 is similar between musicians and nonmusicians, and that the LPC is larger in musicians. (B) Passive listening: Sensory-evoked responses averaged across all mistuning conditions in passive trials separated by group. The topographic maps for each peak show activity at the following latencies: N1, 100 msec; and P2, 180 msec. Electrode Cz is a solid black line, POz is a dotted line, and all other electrodes are gray. Horizontal gray lines show that the amplitude of N1 and P2 is similar between musicians and nonmusicians.

3.3 and 2.3, $p < .05$, but no effect of mistuning was observed for the N1c ($p > .05$). The source of the N1 interaction was an increasing negativity for N1 in musicians but not nonmusicians, whereas the source of the P2 interaction was an increasing negativity for nonmusicians but not musicians. This interaction is likely due to the differing latencies of the ORN between groups and is explained in more detail below. Lastly, the LPC was significantly larger in musicians during active listening, $F(1,26) = 5.4$, $p < .05$, and was not observed in passive trials (see Figure 2). In addition, the effect of mistuning on the LPC was significant, $F(5,130) = 6.81$, $p < .01$. Post hoc tests revealed a smaller LPC at the 2% and the 4% mistuning conditions compared with the tuned condition ($p < .01$ in both cases), whereas no differences in LPC were observed in the 1%, the 8%, and the 16% mistuning conditions compared with the tuned condition ($p > .1$). The mistuning by expertise interaction was not significant for LPC amplitude ($p > .2$).

Object-related Negativity

In both groups, the increase in mistuning was associated with a greater negativity over the 100- to 180-msec time window at fronto-central, $F(5,130) = 16.2$, $p < .01$, and greater positivity at mastoid/cerebellar sites, $F(5, 130) = 16.61$, $p < .01$, consistent with an ORN that was superimposed over the N1 and the P2 waves, with

Downloaded from http://mitpress.mit.edu/jocn/article-pdf/21/8/1483/1937866/jocn.2009.21140.pdf by guest on 18 May 2021

generator(s) in auditory cortices along the superior temporal plane (Figures 3 and 4).

The ANOVA also revealed a significant interaction between musical expertise and mistuning for the ORN recorded at mastoid/cerebellar sites, $F(5,130) = 3.74$, $p < .01$ [linear trend: $F(1,26) = 6.7$, $p < .01$; see Figure 5], with a similar trend for the ORN measured at fronto-central sites, $F(5,130) = 1.7$, $p = .14$ [linear trend: $F(1,26) = 4.93$, $p < .05$]. To gain a better understanding of this interaction, we performed separate ANOVAs for each group. In musicians, pairwise comparisons revealed greater negativity in the 8% and the 16% mistuning conditions compared with the tuned and the 1% conditions ($p < .01$ in all cases). In nonmusicians, only ERPs elicited by the 16% mistuned stimuli differed from those elicited by the tuned stimuli ($p < .05$). This suggests that nonmusicians required greater level of mistuning than musicians to elicit an ORN. In addition, taking into account the polynomial decompositions, these results demonstrate that the ORN is larger in musicians compared with nonmusicians (greater change from tuned to 16% mistuned in musicians compared with nonmusicians at fronto-central: 0.686 versus 0.304 μV and mastoid/cerebellar: 0.888 vs. 0.429 μV). Finally, the interaction between listening condition and mistuning level was not significant nor was the three-way interaction between group, listening condition, and mistuning level ($p > .1$ in all cases). These latter analyses indicate that the ORN was little affected by listening condition in both groups. Finally, the interaction between hemisphere, mistuning, listening condition, and expertise was not significant nor were any lower-order interactions that included hemisphere as a factor at mastoid/cerebellar sites ($p > .1$), indicating no hemispheric asymmetries in ORN amplitude.

To asses the impact of musical expertise on the ORN latency, we measured the peak latency of the difference wave between ERPs elicited by the tuned and those elicited by the 16% mistuned harmonic stimuli. The ORN latency was quantified as the peak activity between 100 and 200 msec poststimulus onset at the midline fronto-central electrode (FCz) in both active and passive listening conditions. The ANOVA, with expertise and listening conditions as factors, yielded a main effect of expertise, with ORN latency being shorter in musicians than in nonmusicians (135 vs. 149 msec), $F(1,26) = 4.28$, $p < .05$. Finally, the main effect of listening condition was not significant nor was the interaction between musical expertise and listening condition, suggesting that the ORN latency is similar in both active and passive listening ($p > .1$ in both cases).

P400

In both groups, the P400 elicited during active listening was slightly right lateralized over the fronto-central scalp

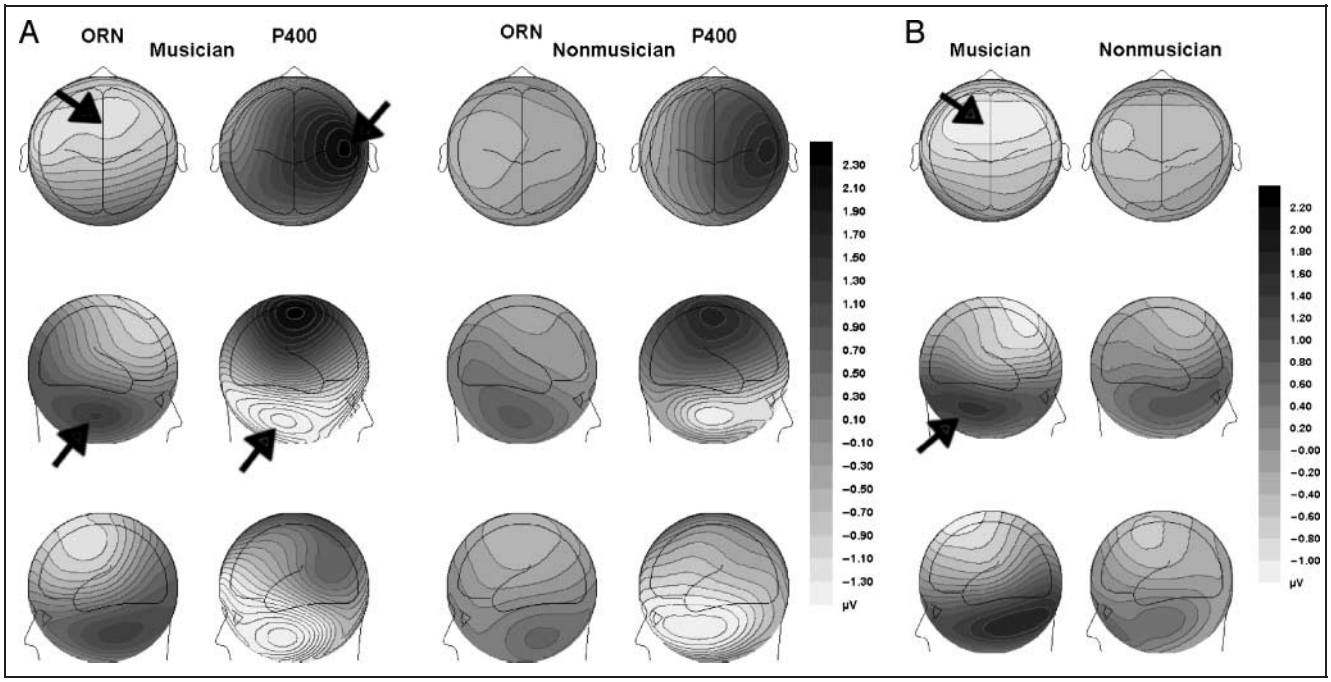

Figure 3. (A) Active listening: Topographic maps of the ORN and the P400 at three angles recorded in during active listening. The ORN contour maps show the peak amplitude for musicians (135 msec) and nonmusicians (149 msec). The P400 contour maps show the mean peak amplitude for musicians (358 msec) and nonmusicians (378 msec). Black arrows in the top row indicate fronto-central ORN and P400 activation; arrows in the middle row show the inversion of the ORN and the P400 at mastoid and cerebellar sites. (B) Passive listening: Topographic maps of the ORN at three angles recorded during passive listening. The latency maps show the ORN amplitude distribution of the ORN at electrode FCz for musicians (135 msec) and nonmusicians (149 msec). Black arrows in the top row indicate fronto-central ORN activation; arrows in the middle row show the inversion of ORN at mastoid and cerebellar sites.

Downloaded from http://mitpress.mit.edu/jocn/article-pdf/21/8/1488/1837086/jocn.2009.21140.pdf by guest on 18 May 2021

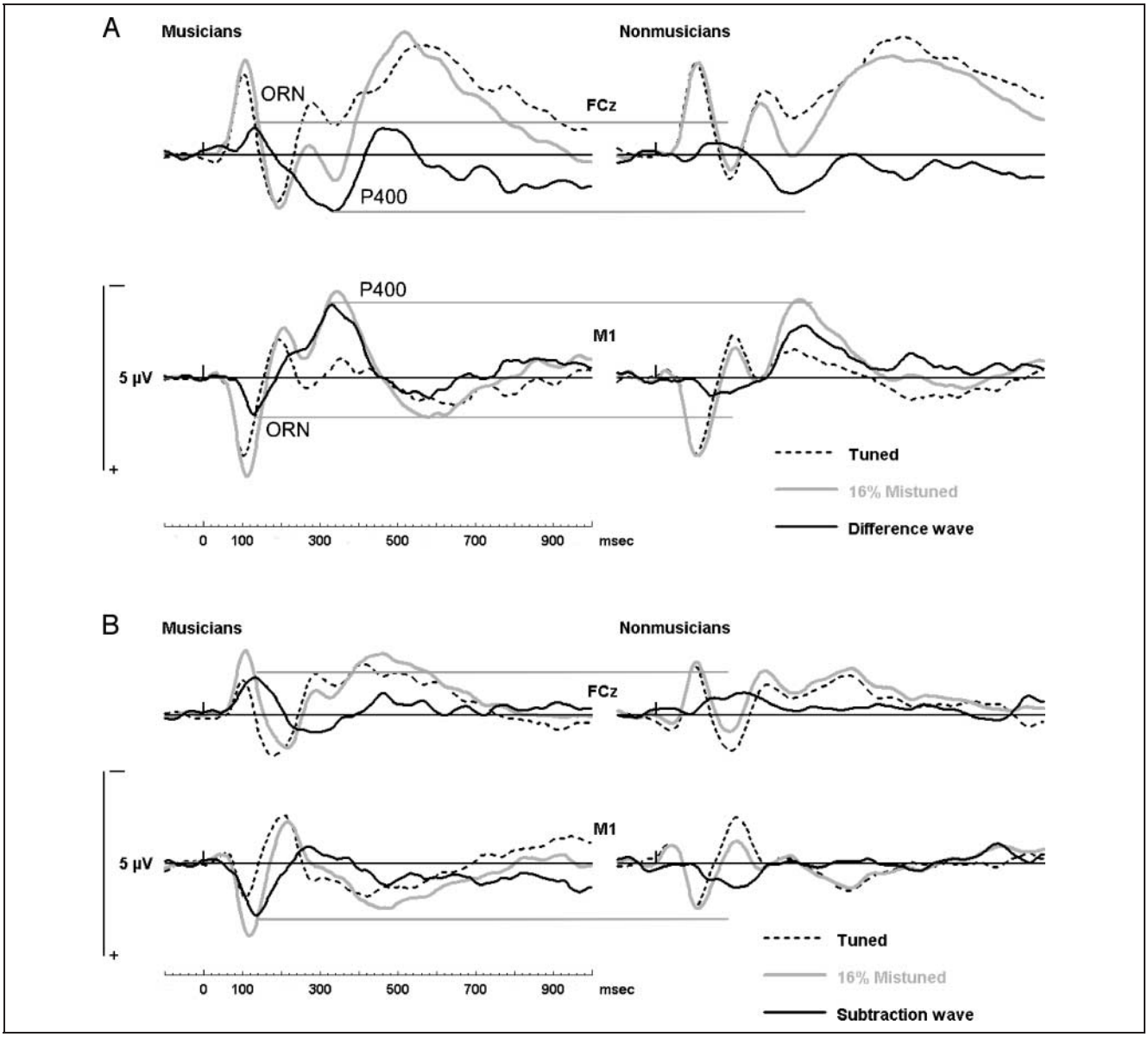

Figure 4. (A) Active listening: The difference between the evoked response in the 16% mistuned stimulus and the tuned stimulus in active trials. The difference wave (in solid black) illustrates the ORN and the P400. Horizontal gray lines from the peak of the ORN and the P400 show the enhancement in musicians. (B) Passive listening: The difference between the evoked response in the 16% mistuned stimulus and the tuned stimulus in passive trials. The difference wave (in solid black) illustrates the ORN. Horizontal gray lines from the peak of the ORN show the enhancement in musicians.

region and inverted in polarity at mastoid/cerebellar sites (Figure 3A). The increase in mistuning was associated with an enhanced positivity over the 300- to 400-msec time window at fronto-central sites, $F(5,130) = 12.52$, $p < .01$, and greater negativity at mastoid/cerebellar sites, $F(5,130) = 13.31$, $p < .01$, consistent with a P400 with generator(s) in auditory cortices along the superior temporal plane (Figures 3A and 4A).

More importantly, the ANOVA on the mean amplitude over the 300- to the 400-msec interval yielded an interaction between musical expertise and mistuning at mastoid/cerebellar sites, $F(5,130) = 2.50$, $p < .05$ [quadratic trend, $F(1,26) = 8.37$, $p < .01$], and fronto-central sites, $F(5, 130) = 2.40$, $p < .05$ [quadratic trend, $F(1,26) = 1.55$, $p < .01$]. To gain a better understanding of this interaction, we performed separate ANOVAs for each group. In musicians, pairwise comparisons revealed greater positivity in the 8% and the 16% mistuning conditions compared with the tuned and the 1% conditions ($p < .05$ in all cases). In nonmusicians, ERPs elicited by the 8% and the 16% mistuned stimuli differed from those elicited by only the tuned stimuli ($p < .05$ in both cases). This suggests that both groups required similar levels of mistuning to elicit a P400. Taking into account the polynomial decompositions, the P400 was elicited with similar levels of mistuning but was larger in musicians (greater change from tuned to 16% mistuned in musicians compared with nonmusicians at fronto-central 1.02 vs. 0.77 μV

Downloaded from http://mitpress.mit.edu/jocn/article-pdf/21/8/1483/1937866/jocn.2009.21140.pdf by guest on 18 May 2021

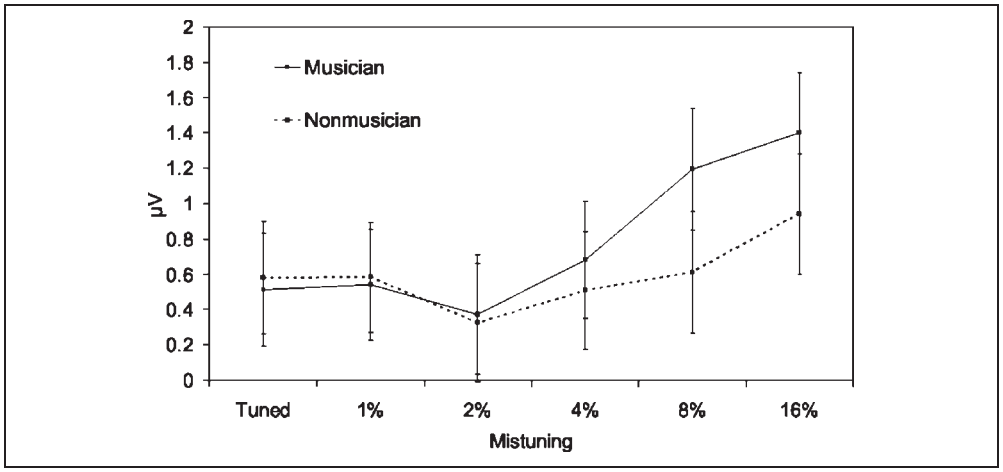

Figure 5. ORN 100–180 msec: Mean amplitude of the evoked response averaged across four mastoid electrodes, from 100 to 180 msec poststimulus onset, as a function of mistuning (error bars = 1 *SE*).

and mastoid/cerebellar 1.29 vs. 0.78 μV). Finally, the interaction between hemisphere, mistuning, and musical expertise was not significant nor were any lower-order interactions that included hemisphere as a factor ($p > .1$). There was a significant main effect of hemisphere at mastoid/cerebellar sites, $F(1,26) = 10.35$, $p < .01$, indicating greater activity (not P400 because P400 requires a mistuning effect) recorded over the right hemisphere.

The P400 latency was defined as the largest peak on the difference wave (ERPs to tuned stimuli minus ERPs elicited by the 16% mistuned stimuli) at electrodes C2 and C4 during the 250- to 450-msec interval. The latency of the P400 was slightly shorter in musicians compared with nonmusicians (358 vs. 378 msec); however, this effect was not statistically reliable ($p > .1$).

DISCUSSION

The purpose of this study was to examine the influence of long-term training on concurrent sound segregation. We found that musicians were more likely to identify a mistuned harmonic as a distinct auditory object compared with nonmusicians. This was paralleled by larger amplitude and earlier ORN waves and larger P400 waves. Our behavioral and electrophysiological data demonstrate that musicians have enhanced ability to partition the incoming acoustic wave based on harmonic relations. More importantly, these results cannot easily be accounted for by models of auditory scene analysis that postulate that low-level processes occur independently of listeners' experience. Instead, the findings support the more contemporary idea that long-term training can alter even primitive perceptual functions (see Wong et al., 2007; Koelsch, Schroger, & Tervaniemi, 1999; Beauvois & Meddis, 1997).

The earlier and enhanced ORN amplitude in musicians likely reflects greater abilities in the primitive processing of periodicity cues. Studies measuring the mismatch negativity (MMN) wave, an ERP component thought to index a change detection process (e.g., Picton et al., 2000; Näätänen, Gaillard, & Mantysalo, 1978), have shown enhancements to the MMN in musicians across numerous domains, including violations of periodicity (Koelsch et al., 1999), violations of melodic contour and interval structure (Fujioka, Trainor, Ross, Kakigi, & Pantev, 2004, 2005), and violations of temporal structure (Russeler, Altenmuller, Nager, Kohlmetz, & Munte, 2001). Interestingly, Koelsch et al. (1999) found that when the same components of the harmonic series were presented in isolation, the deviant mistuned tone evoked a comparable MMN in both musicians and nonmusicians; however, when the same deviant sound was presented as part of a chord, musicians had a larger MMN and were able to identify the deviant chord more consistently. Therefore, although both musicians and nonmusicians can detect differences in frequency, musicians have an advantage when dealing with concurrently occurring sounds and detecting violations of periodicity.

Detection of periodic (harmonic) violations must precede or coincide with concurrent sound segregation because without detection, perception of a second auditory object would be impossible. Although musical training did not alter the amount of mistuning required to perceive a second auditory object (2–4% in both groups), musicians were more consistent in their perceptions, which suggests that as a result of musical training, harmonic violations are more easily detected by musicians. The increased ability of musicians to detect mistuning in a complex sound allows for more consistent sound segregation.

Koelsch et al. (1999) observed musician-related enhancements at identifying mistuning in a complex sound. The Koelsch et al. study used pure tones arranged as chords, which isolated the harmonic relations found in music, without using timbres of musical instruments, much like the current study used mistuned harmonics to investigate sound segregation without using stimuli with musical timbres. Isolating low-level perceptual functions (from the effect of timbre) is paramount to drawing conclusions about low-level scene analysis functions

Downloaded from http://direct.mit.edu/jocn/article-pdf/21/8/1488/1937086/jocn.2009.21140.pdf by guest on 18 May 2021

because previous research has shown enhanced amplitude for the N1 (Pantev et al., 1998, 2001), N1c (Shahin et al., 2003), and P2 (Shahin et al., 2003, 2005) in musicians when presented with stimuli of musical timbre. The enhancements to the N1, the N1c, and the P2 in musicians are typically observed for musical sounds, especially for those that are similar to the instrument of training (e.g., piano tone for pianist, trumpet sounds for trumpeter). The expertise-related differences in sensory-evoked responses are typically small or even nonexistent when musicians and nonmusicians are presented with pure tones (see Shahin et al., 2003, 2005).

It is important to acknowledge the cortical source of the myriad enhancements observed in musicians. Long-latency auditory-evoked responses (i.e., N1, N1c, and P2) are thought to originate at various points along the superior temporal plane (see Scherg, Vajsar, & Picton, 1999), and therefore enhancements to these waveforms were thought to be due to cortical plasticity. Emerging evidence suggests that the plasticity goes even deeper and may be at the level of the brainstem (Wong et al., 2007). Taking this new data into account, one could hypothesize that enhancements to long latency auditory-evoked responses are due to a stronger signal coming in from the brain stem. In terms of the present study, the ORN enhancements could be due to enhanced frequency coding at precortical stages of the auditory pathway, as a reliable ORN emerges with less mistuning in musicians compared with nonmusicians. The data from the present study cannot support or refute this hypothesis, and further study is warranted.

In the present study, cortical representations of harmonic complexes (as indexed by N1, N1c, and P2 waves) were similar in both musicians and nonmusicians. Group differences were only observed in ERP components related to the perception of simultaneous sounds. Harmonic complexes are not domain specific to music; thus, the lack of effects on the N1, the N1c, and the P2 waves were to be expected. Musicians do, however, segregate simultaneous sounds as part of their training. Performers in a large group must be able to segregate instruments from one another; even practicing alone requires the musician to segregate the sounds of his or her instrument from environmental noise. Some of this segregation is probably based on harmonicity, which may be why musicians demonstrate enhanced concurrent sound processing.

The use of harmonicity as a cue for auditory scene analysis in a musical setting also explains the enhancement to the LPC. The LPC has been described as an index of the decision-making process about an incoming sound stimulus (Starr & Don, 1988). The data in the current study support this explanation because the LPC was smallest in conditions where the decision about the harmonic complex was difficult (2–4% mistuning) for both groups. This may be related to the increased variance in behavioral performance in the 2% and the 4% mistuning conditions, indicating that LPC amplitude might be related to the confidence in behavioral responses. In addition, a larger LPC was observed in musicians. Previous research demonstrated increased LPC activity in musicians when making decisions about terminal note congruity (Besson & Faita, 1995). The enhanced LPC in musicians in the current study may be due to the salience of periodicity and violations of periodicity for musicians. For a performing musician, different cues would require different behavioral responses. For example, a violinist in a group may determine that she is slightly out of tune with the rest of the group and adjust her fingering accordingly. For the lay person, slight harmonic violations are not normally important. This alternative explanation suggests that the change in the LPC observed in musicians is due to cortical enhancements related to harmonic detection and related actions.

Despite the evidence for the effect of musical expertise on primitive auditory scene analysis, some alternative explanations should be considered. One possibility is that musicians were better at focusing their attention to the frequency region of the mistuned harmonic. In the present study, musicians may have realized that it was always the second harmonic that was mistuned and used this information to focus their attention to the frequency of the mistuned harmonic. Although the bulk of research suggests that the ORN indexes an attention-independent process (Alain, 2007), there is some evidence that under certain circumstances (i.e., when the mistuned harmonic is predictable) the ORN amplitude may be enhanced by attention (see Experiment 1 by Alain et al., 2001). Hence, the enhancements observed in the ORN of musicians could be due to a greater allocation of attention to the frequency region of the mistuned harmonic. The data, however, does not support this view. Nonsignificant interactions between mistuning and listening condition and between mistuning, listening condition, and musical training indicate that the observed effects were consistent in both passive and active listening. The ORN was enhanced in musicians compared with nonmusicians by similar amounts in both listening conditions.

Another possible explanation for our findings is that in the present study we used a strict selection criterion for nonmusicians, excluding participants with intermediate levels of musical training. By using a strict criterion for selecting nonmusicians, we may have selected individuals who have poor auditory processing abilities in general. Individuals with poor auditory abilities may not have been detected using pure tone thresholds as the sole screening procedure. Future research should consider a more comprehensive assessment of auditory abilities when comparing musicians and nonmusicians. Poor auditory processing abilities could explain why the ORN of the nonmusicians was much smaller compared with the ORN observed in previous studies (where musical training was not a criterion). Similarly, in the

Downloaded from http://mitpress.mit.edu/jocn/article-pdf/21/8/1483/1937186/jocn.2009.21140.pdf by guest on 18 May 2021

present study, we aimed to select a group of highly trained musicians who may have enhanced auditory processing abilities. Thus, our screening method may have created two groups at opposite ends of the spectrum in term of auditory abilities.

Conclusion

The findings of the current study support the hypothesis that musical training enhances concurrent sound segregation. Music perception is governed by the same primitive auditory scene processes as all other auditory perception. Bregman (1990) points out that ''the primitive processes of auditory organization work in the same way whether they are tested by studying simplified sounds in the laboratory or by examining examples in the world of music'' (p. 528). If we apply this theory to the current data, we can conclude that musical training engenders general enhancements to concurrent sound segregation, regardless of stimulus type.

The process of concurrent sound segregation is different in expert musicians. Musicians are better at identifying concurrently occurring sounds, and this is paralleled by neural change. This positive change in musicians is probably due to experience in dealing with chords and other harmonic (and inharmonic) relations found in music. Enhancements to concurrent sound segregation and related neural activity suggest that primitive auditory scene abilities are improved by long-term musical training.

Acknowledgments

The research was supported by grants from the Canadian Institutes of Health Research and the Natural Sciences and Engineering Research Council of Canada. Special thanks to Dr. Takako Fujioka, Dr. Ivan Zendel, Patricia Van Roon, and two anonymous reviewers for constructive comments on earlier versions of this manuscript.

Reprint requests should be sent to Claude Alain, Rotman Research Institute, Baycrest Centre for Geriatric Care, 3560 Bathurst Street, Toronto, Ontario, Canada M6A 2E1, or via e-mail: calain@rotman-baycrest.on.ca.

Note

1. The N1 wave refers to a deflection in the auditory ERPs that peaks at about 100 msec after sound onset and is largest over the fronto-central scalp region. It is followed by an N1c, which is a smaller negative wave over the right and the left temporal sites and a P2 wave that peaks at about 180 after sound and is maximal over the central scalp region. For a more detailed review of long-latency human auditory-evoked potentials, see Crowley and Colrain (2004), Scherg et al. (1999), Starr and Don (1988), and Näätänen and Picton (1987).

REFERENCES

Alain, C. (2007). Breaking the wave: Effects of attention and learning on concurrent sound perception. *Hearing Research, 229,* 225–236.

Alain, C., Arnott, S. R., & Picton, T. W. (2001). Bottom–up and top–down influences on auditory scene analysis: Evidence from ERPs. *Journal of Experimental Psychology, 27,* 1072–1089.

Alain, C., & Izenberg, A. (2003). Effects of attentional load on auditory scene analysis. *Journal of Cognitive Neuroscience, 15,* 1063–1073.

Alain, C., Schuler, B. M., & McDonald, K. L. (2002). Neural activity associated with distinguishing concurrent auditory objects. *Journal of the Acoustical Society of America, 111,* 990–995.

Alain, C., & Snyder, J. S. (2008). Age-related differences in auditory evoked responses during rapid perceptual learning. *Clinical Neurophysiology, 119,* 356–366.

Alain, C., Snyder, J. S., He, Y., & Reinke, K. (2007). Changes in auditory cortex parallel rapid perceptual learning. *Cerebral Cortex, 17,* 1074–1084.

Beauvois, M. W., & Meddis, R. (1997). Time decay of auditory stream biasing. *Perception and Psychophysics, 59,* 81–86.

Besson, M., & Faita, F. (1995). An ERP study of musical expectancy: Comparison of musicians with nonmusicians. *Journal of Experimental Psychology, 21,* 1278–1296.

Bey, C., & McAdams, S. (2002). Schema-based processing in auditory scene analysis. *Perception and Psychophysics, 64,* 844–854.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: Age, sleep & modality. *Clinical Neurophysiology, 115,* 732–744.

Dowling, W. J. (1973). The perception of interleaved melodies. *Cognitive Psychology, 5,* 322–337.

Fishman, Y. I., Volkov, I. O., Noh, M. D., Garell, P. C., Bakken, H., Arezzo, J. C., et al. (2001). Consonance and dissonance of musical chords: Neural correlates in auditory cortex of monkeys and humans. *Journal of Neurophysiology, 86,* 2761–2788.

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., & Pantev, C. (2004). Musical training enhanced automatic encoding of melodic contour and interval structure. *Journal of Cognitive Neuroscience, 16,* 1010–1021.

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., & Pantev, C. (2005). Automatic encoding of polyphonic melodies in musicians and non-musicians. *Journal of Cognitive Neuroscience, 17,* 1578–1592.

Hafter, E. R., Schlauch, R. S., & Tang, J. (1993). Attending to auditory filters that were not stimulated directly. *Journal of the Acoustical Society of America, 94,* 743–747.

Koelsch, S., Schroger, E., & Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *NeuroReport, 10,* 1309–1313.

Moore, B. C., Glasberg, B. R., & Peters, R. W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *Journal of the Acoustical Society of America, 80,* 479–483.

Näätänen, R., Gaillard, A. W. K., & Mantysalo, S. (1978). Early selective attention effect on evoked potential reinterpreted. *Acta Psychologica, 42,* 313–329.

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology, 24,* 375–425.

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature, 392,* 811–814.

Pantev, C., Roberts, L. E., Schultz, M., Engelien, A., & Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *NeuroReport, 12,* 169–174.

Downloaded from http://mitpress.mit.edu/jocn/article-pdf/21/8/1488/1937186/jocn.2009.21140.pdf by guest on 18 May 2021

Picton, T. W., van Roon, P., Armilio, M. L., Berg, P., Ille, N., & Scherg, M. (2000). The correction of ocular artifacts: A topographic perspective. *Clinical Neurophysiology, 111,* 53–65.

Russeler, J., Altenmuller, E., Nager, W., Kohlmetz, C., & Munte, T. F. (2001). Event related brain potentials to sound omissions differ in musicians and non-musicians. *Neuroscience Letters, 308,* 33–36.

Reinke, K. S., He, Y., Wang, C., & Alain, C. (2003). Perceptual learning modulates sensory evoked response during vowel segregation. *Cognitive Brain Research, 17,* 781–791.

Scherg, M., Vajsar, J., & Picton, T. W. (1999). A source analysis of the late human auditory evoked potentials. *Journal of Cognitive Neuroscience, 1,* 336–355.

Schlauch, R. S., & Hafter, E. R. (1991). Listening bandwidths and frequency uncertainty in pure-tone signal detection. *Journal of the Acoustical Society of America, 90,* 1332–1339.

Shahin, A., Bosnyak, D. J., Trainor, L. J., & Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *Journal of Neuroscience, 23,* 5545–5552.

Shahin, A., Roberts, L. E., Pantev, C., Trainor, L. J., & Ross, B. (2005). Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. *NeuroReport, 16,* 1781–1785.

Sinex, D. G., Guzik, H., Li, H., & Sabes, J. H. (2003). Responses of auditory nerve fibers to harmonic and mistuned complex tones. *Hearing Research, 182,* 130–139.

Sinex, D. G., Sabes, J. H., & Li, H. (2002). Responses of inferior colliculus neurons to harmonic and mistuned complex tones. *Hearing Research, 168,* 150–162.

Starr, A., & Don, M. (1988). Brain potentials evoked by acoustic stimuli. In T. W. Picton (Ed.), *Human event-related potentials: EEG handbook* (Vol. 3, pp. 97–157). Amsterdam: Elsevier Science Publishers.

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

Downloaded from http://direct.mit.edu/jocn/article-pdf/21/8/1483/1937086/jocn.2009.21140.pdf by guest on 18 May 2021