Published in final edited form as:
Eur J Neurosci. 2013 April ; 37(8): 1295–1307. doi:10.1111/ejn.12110.

Musicians Show General Enhancement of Complex Sound Encoding and Better Inhibition of Irrelevant Auditory Change in Music: An ERP Study

Natalya Kaganovich[*,1,2], **Jihyun Kim**[2], **Caryn Herring**[1], **Jennifer Schumaker**[1], **Megan MacPherson**[1], and **Christine Weber-Fox**[1]

[1]Department of Speech, Language, and Hearing Sciences, Purdue University, 500 Oval Drive West Lafayette, IN 47907-2038

[2]Department of Psychological Sciences, Purdue University, 703 Third Street, West Lafayette, IN 47907-2038

Abstract

Using electrophysiology, we have examined two questions in relation to musical training – namely, whether it enhances sensory encoding of the human voice and whether it improves the ability to ignore irrelevant auditory change. Participants performed an auditory distraction task, in which they identified each sound as either short (350 ms) or long (550 ms) and ignored a change in sounds' timbre. Sounds consisted of a male and a female voice saying a neutral sound [a], and of a cello and a French Horn playing an F3 note. In some blocks, musical sounds occurred on 80% of trials, while voice sounds on 20% of trials. In other blocks, the reverse was true. Participants heard naturally recorded sounds in half of experimental blocks and their spectrally-rotated versions in the other half. Regarding voice perception, we found that musicians had a larger N1 ERP component not only to vocal sounds but also to their never before heard spectrally-rotated versions. We, therefore, conclude that musical training is associated with a general improvement in the early neural encoding of complex sounds. Regarding the ability to ignore irrelevant auditory change, musicians' accuracy tended to suffer less from the change in sounds' timbre, especially when deviants were musical notes. This behavioral finding was accompanied by a marginally larger re-orienting negativity in musicians, suggesting that their advantage may lie in a more efficient disengagement of attention from the distracting auditory dimension.

Keywords

musical training; auditory perception in musicians; timbre perception; auditory change; auditory distraction

1. Introduction

This study has examined two questions in relation to musical training – namely, whether it enhances sensory encoding of the human voice due to the latter's perceptual similarity to musical sounds and whether it improves the ability to ignore irrelevant auditory change. Previous research has shown that musical training leads to enhancement in the sensory encoding of musical sounds as revealed by the increased amplitude of the N1 and P2 event-

[*]Corresponding author: Department of Speech, Language, and Hearing Sciences Purdue University 500 Oval Drive West Lafayette IN 47907-2038 Phone (765)494-4233 Fax (765)494-0771 kaganovi@purdue.edu.

The authors report no conflict of interest.

related potential (ERP) components in musicians compared to non-musicians (e.g., Pantev *et al.*, 1998; Shahin *et al.*, 2003; Shahin *et al.*, 2004; Fujioka *et al.*, 2006). Such enhancement is greater for the instrument of training (e.g., Pantev *et al.*, 2001), with some of its aspects already evident in brain stem recordings (Strait *et al.*, 2012). We asked whether musicians' superiority in the early processing of musical timbre may extend to the perceptually similar timbre of the human voice. Although acoustic correlates of musical and vocal timbre have been studied largely independently from each other, in both cases the perceived timbre is due to a combination of multiple temporal and spectral properties of sound (Handel, 1989; McAdams *et al.*, 1995; Kreiman, 1997; Caclin *et al.*, 2005). Furthermore, neuropsychological and brain imaging studies point to similarities in the brain areas involved in vocal and musical timbre processing (Peretz *et al.*, 1994; Samson & Zatorre, 1994; Peretz *et al.*, 1997; Samson *et al.*, 2002; von Kriegstein *et al.*, 2003; Halpern *et al.*, 2004), suggesting that the perception of both timbres may rely on similar neural and cognitive processes. We, therefore, hypothesized that musical training, and more specifically, a systematic exposure to the timbre of musical instruments, may be associated with an enhanced sensory encoding of the human voice.

A few recent studies suggest that musical training may also lead to improvement in attentional control (Trainor *et al.*, 2009; Strait & Kraus, 2011). However, the concept of attention covers a large range of abilities, and existing research is only beginning to evaluate how musical training may differentially affect its various facets. One aspect of attention that may be influenced by musical training, but which has not as yet been investigated, is the ability to ignore irrelevant auditory change. More specifically, musical training requires one to focus on some aspects of musical signal while ignoring others, as for example in identifying the same note across multiple instruments or across multiple octaves. We, therefore, hypothesized that musical training may be associated with a better ability to screen out those auditory changes that are not relevant for the task at hand.

We tested both hypotheses by employing a version of the auditory distraction paradigm (Schröger & Wolff, 1998; 2000), in which participants categorized sounds by their length and ignored task-irrelevant changes in the sounds' timbre (vocal vs. musical). We used the N1 ERP component as a measure of early auditory processing and the P3a, P3b, and the re-orienting negativity ERP components as measures of distraction and a successful return to the categorization task.

2. Method

2.1 Participants

Behavioral and ERP data were collected from 19 musicians (11 female) and 17 non-musicians (10 female). All participants were students at Purdue University at the time of testing and participated either for course credit or for payment. The participants' age was 20.2 years for musicians and 20 years for non-musicians, on average (group, $F(1,35)<1$). All participants were free of neurological disorders, based on self-report, passed a hearing screening at a level of 20 dB HL at 500, 1000, 2000, 3000, and 4000 Hz, reported to have normal or corrected-to-normal vision, and were not taking medications that may affect brain function (such as anti-depressants) at the time of study. All gave their written consent to participate in the experiment, which was approved by the Institutional Review Board of Purdue University. All study procedures conformed with the The Code of Ethics of the World Medical Association (Declaration of Helsinki) (1964).

The group of musically trained participants consisted of amateur musicians. To be included in this group, a participant had to meet the following criteria. (1) The onset of musical training had to occur prior to the age of 12 (the average onset was 7.5 years of age; range

3-11). (2) The duration of musical training had to be at least 5 years (the average duration was 9.3 years; range 5-15 years). Musical training could include either formal musical instruction or participation in musical groups (such as a high school marching band, for example). (3) And lastly, an individual had to be a member of a musical organization or group either currently or in the past. Such groups ranged from middle and high school concert and marching bands to Purdue University musical groups. These criteria were designed to select individuals who had significantly more musical training than an average non-musician while not reaching the level of professional musicians. All musicians received training for more than one instrument. Four listed voice as one of their expertise areas, but none of the musicians trained in voice exclusively. Additionally, none of the musicians listed either a cello or a French Horn (whose sounds were used as stimuli in the current study) as their primary or secondary instruments of training.

2.2 Stimuli

Stimuli consisted of two sound categories – human voices and musical instruments. The voice category contained natural recordings of a male and a female voice saying a neutral sound [a]. The musical instruments category contained natural recordings of a cello and a French Horn playing an F3 note. Both types of stimuli were equated in frequency (174 Hz), which remained constant for the duration of the sound. This was achieved by asking speakers to match the pitch of a pre-recorded tone. Speakers were successful within a few Hz. The remaining frequency difference was corrected in Praat 5.1. (Boersma & Weenink, 2011). Each sound had two durations – 350 ms and 550 ms. The short duration sound was created by reducing the length of all parts of the long duration sound in Praat 5.1. Spectrally-rotated versions of all sounds were generated by rotating their frequencies around 2000 Hz (MATLAB R2010b). Spectrally-rotated sounds retained their complexity, pitch, periodicity, and the overall temporal envelope as can be seen in their waveforms and spectrograms shown in Figure 1. However, the timbre of original sounds was completely altered and no longer resembled any of the naturally produced sounds (Blesser, 1972).

To equate for perceptual loudness, the male and female voice stimuli were presented at 70 dB SPL, while the cello and the French Horn stimuli at 73 and 74 dB SPL respectively. These values were selected during a pilot study in which participants were asked to judge whether the four sounds (male voice, female voice, cello, and French Horn) sounded equally loudly. The intensity of spectrally-rotated sounds was matched with that of their natural counterparts. Sounds were presented in free field via a single speaker (SONY) located approximately four feet in front of a participant and directly above the computer monitor that displayed instructions and a hair-cross point for eye fixation.

2.3 Design and Experimental Measures

We used the auditory distraction paradigm developed by Schröger and colleagues (Schröger & Wolff, 1998; 2000). The study had 2 conditions, with 4 blocks in each. The first condition consisted of natural sounds (NAT); and the second condition of spectrally-rotated sounds (ROT). Within each condition, two blocks contained musical sounds as standards (P=0.8) and vocal sounds as deviants (P=0.2), while the reverse was true for the other two blocks. In each block, standards always consisted of just one token of one sound category (for example, just a female voice), while deviants consisted of both tokens of the opposite sound category (for example, a cello and a French Horn). Both tokens of the deviant sounds occurred equiprobably (P=0.1 each). Both standards and deviants were of two durations – 350 and 550 ms, with each duration occurring equiprobably (P=0.5). The 200 ms difference between short and long sounds used in the current study is similar to that used in previous studies employing the auditory distraction paradigm (e.g., Schröger & Wolff, 2000; Mager *et al.*, 2005). Each block consisted of 200 trials (160 standards and 40 deviants). The interstimulus interval varied randomly between 1.6 and 2 seconds. The ROT condition was identical to the NAT condition, with the only difference being that spectrally-rotated versions of voices and musical sounds were used throughout. Table 1 lists all conditions and blocks of the study. Figure 2 details the structure of a single block.

The eight blocks of the experiment were presented in a Latin square design, lasting approximately 6 minutes each. Participants were instructed to press one response button for short sounds and another for long sounds. Three examples of short and three of long sounds present in each block were always played at the beginning of a block to familiarize participants with the sounds they were instructed to categorize. Hand to response button mapping was counterbalanced across participants. Importantly, unlike in odd-ball paradigms, responses were provided for all sounds (i.e. standards and deviants).

The N1 ERP component elicited by the onset of auditory stimuli was evaluated in order to compare the two groups' early neural processing of musical and vocal sounds. N1 is a measure of early sensory encoding of the physical properties of sound, such as frequency, complexity, and intensity, among others (Näätänen & Picton, 1987). Importantly, previous ERP research has demonstrated the influence of musical training on the amplitude of N1 and its N1c subcomponent. For example, it has been shown to be enhanced in musicians in response to both musical notes and pure tones, with greater enhancement for the sounds of the instrument of training (e.g., Pantev *et al.*, 1998; Pantev *et al.*, 2001; Shahin *et al.*, 2003; Baumann *et al.*, 2008). We, therefore, predicted that musicians would exhibit a larger N1 component to musical sounds. We also speculated that given the similarity of vocal and musical timbres and their underlying acoustic properties, musicians might also show an enhanced N1 to voices.

Additionally, we have evaluated several behavioral and electrophysiological measures associated with distraction. A change in timbre (for example, from a musical instrument to a voice or vice versa) within each block was irrelevant for the duration discrimination task. However, because each sound contained both task-relevant (duration) and task-irrelevant (timbre) information, timbre change was expected to lead to a distraction. Such distraction and the following successful return to the task have been associated with both behavioral and electrophysiological indices (Schröger & Wolff, 1998; 2000). Behaviorally, we expected longer reaction times (RT) and/or lower accuracy (ACC) for deviant sounds, which signaled the timbre change. Electrophysiologically, a distraction is typically manifested as a fronto-central P3a component to deviant sounds, indicating the distraction process itself (e.g., Pollich, 2003). It is followed by a central-parietal P3b component, reflecting a working memory update in response to the noticed change (Donchin, 1981; Donchin & Coles, 1988; Picton, 1992; Polich, 2007). Lastly, these two components are followed by the re-orienting negativity (RON) thought to reflect a successful return to the task at hand (Schröger & Wolff, 1998; 2000; Horváth *et al.*, 2008; Horváth *et al.*, 2011), with larger amplitude of RON being indicative of a more successful disengagement of attention from the distracting dimension. Although the exact components elicited by different versions of the RON paradigm may differ somewhat from study to study, the sequence of expected components described above is typical for this paradigm and has been reported in previous studies (e.g, Schröger & Wolff, 2000; Horváth *et al.*, 2008; Wronka *et al.*, 2012) . In order to keep the length of the study within reasonable limits, we did not include a passive listening condition (which was used in some of the earlier RON studies) in our design. Passive listening is the choice paradigm for eliciting the mismatch negativity (MMN) ERP component, which is thought to index the automatic detection of auditory change (e.g., Näätänen, 1995; Näätänen & Alho, 1995; Näätänen, 2001). Although MMN may be elicited during participants' active engagement in a task, its evaluation under these circumstances is often made difficult by overlapping task-specific ERP components (e.g., Sussman, 2007). We, therefore, chose not to evaluate MMN differences between the two groups. Additionally, the P2 ERP component was elicited by standards but was overlapped by the closely following P3a in deviants. Due to this overlap and a significant amplitude enhancement of the N1 component in musicians, which appeared to have spread to the temporal window of the adjacent P2, the P2 component was not analyzed.

We expected that musicians would be less affected by irrelevant timbre change, and that, as a group, they would show a smaller RT increase and a smaller ACC drop in response to deviant sounds. We further expected that this superior behavioral performance might be reflected in ERPs as smaller P3a and P3b components compared to non-musicians. We also expected that musicians would show a larger RON as an indicator of a greater success at returning to the duration discrimination task after a distraction took place. In sum, RT, ACC, P3a, P3b, and RON were our main measures for evaluating the group difference between musicians and non-musicians in the ability to ignore irrelevant auditory change.

Lastly, we wanted to understand to what extent expected advantages in the musicians group can generalize to completely novel sounds by examining ERPs elicited not only by naturally recorded sounds but also by their ROT versions. While ROT sounds retained some of the acoustic properties (such as complexity, pitch, periodicity, and temporal envelope) of NAT sounds, their original timbre was completely unrecognizable. We hypothesized that if moderate musical training leads to benefits that are tightly coupled with the specific timbres to which a musician is exposed, then we should see the expected benefits in the NAT condition but not in the ROT condition. However, if moderate musical training is associated with a more general enhancement of complex sound encoding and cognitive control, musicians may show advantages in both conditions.

In addition to the main task described above, all participants were administered the Melody part of the *Music Aptitude Profile* (Gordon, 2001) to obtain a more objective measure of their musical ability. They also filled out a detailed questionnaire on their musical training and experience.

2.4 Event-Related Potentials Recordings

Electrical activity was recorded from the scalp using 32 Ag-Cl electrodes secured in an elastic cap (Quik-cap). Electrodes were positioned over homologous locations across the two hemispheres according to the criteria of the International 10-20 system (American Electroencephalographic Society, 1994). The specific locations were: midline sites FZ, FCZ, CZ, CPZ, PZ, OZ; mid-lateral sites: FP1/FP2, F3/F4, FC3/FC4, C3/C4, CP3/CP4, P3/P4, O1/O2; and lateral sites: F7/F8, FT7/FT8, T7/T8, TP7/TP8, P7/P8; and left and right mastoids. Electroencephalographic activity was referenced to the left mastoid and re-referenced offline to the average of the left and right mastoids (Luck, 2005). The electro-oculograms were bipolar recordings via electrodes placed over the right and the left outer canthi (horizontal eye movement) and left inferior and superior orbital ridge (vertical eye movement). The electrical signals were amplified between 0.1 and 100Hz and digitized online (Neuroscan 4.2) at the rate of 500 samples per second.

Individual EEG records were visually inspected to exclude trials containing excessive muscular and other non-ocular artifacts. Ocular artifacts were corrected by applying a spatial filter (EMSE Data Editor, Source Signal Imaging, Inc.). ERPs were epoched starting at 200 ms pre-stimulus and ending at 900 ms post-stimulus onset. The 200 ms prior to the recording onset served as a baseline. Peak amplitudes of components were computed as the maximum (or the minimum) voltage within a certain time window, which was both preceded and followed by a smaller voltage measurement. Peak latencies were computed relative to the onset of a stimulus (0 ms). Mean amplitudes were calculated as mean voltages within a specified temporal window. Peak amplitude and peak latency were used to evaluate the N1 ERP component. Mean amplitude was used to evaluate all other ERP components to avoid the effect of latency jitter (Luck, 2005).

2.5 Data analysis

Both behavioral and ERP analyses compared responses elicited by acoustically identical sounds when such sounds functioned as standards and as deviants. Thus responses to voice deviants were compared with responses to voice standards and responses to music deviants were compared with responses to music standards, etc. RT and ACC for standards and deviants, as well as the difference in RT and ACC between standards and deviants were calculated. These measures were pooled across every two blocks in which the same sound category (i.e. voice or musical instrument) was used as deviants (see Table 1). A preliminary analysis of RT and ACC revealed no group by sound duration interaction (*RT*: NAT, $F(1,34)<1$; ROT, $F(1,34)=1.568$, $p=0.219$; *ACC*: NAT, $F(1,34)<1$; ROT, $F(1,34)=1.782$, $p=0.191$); therefore, data were also pooled over the short and long durations of the same sound. For example, long and short male and female voice trials were averaged together to represent “voice standards,” and short and long cello and French Horn sounds were averaged together to represent “music deviants.”

Analysis of ERP data was parallel to that of behavioral measures. For each electrode site, ERP trials were averaged separately for standards and deviants across each two blocks in which the same sound type was used as a deviant. Because the pattern of group differences was not affected by the length of stimuli and to increase the signal-to-noise ratio, ERPs elicited by short and long sounds were averaged together for each stimulus type (i.e. standard and deviant). This approach to data analysis is similar to that used by Schröger and colleagues (Schröger & Wolff, 2000). Although sound length was not included as a variable in data analysis due to too few ERP trials available for each length, examples of ERP responses to short and long versions of the same sound are included in all ERP figures to demonstrate that the pattern of responses did not differ significantly between the two lengths. Time windows and sites used for each component’s analyses were selected in agreement with the official guidelines for recording human event-related potentials (Picton *et al.*, 2000) and current practices in the field (Luck, 2005). Selection of sites was based on the grand average waveforms and a typical distribution of any given component. Table 2 lists time windows and electrode sites used for each component’s analysis. Acoustic differences between NAT and ROT sounds resulted in a slight difference in the latency of the same components in the NAT and ROT conditions. In order to optimize component measurements, time windows for selecting individual components were positioned over slightly different portions of ERP recordings in the NAT and ROT conditions as indicated in Table 2; however, the length of the time window over which any given component was measured (e.g., N1) was kept constant across the board.

Because absolute values of the P3a, P3b, and RON components of the deviant waveforms carry little significance without a comparison with standard waveforms, we first measured the mean amplitude of these components as elicited by both the deviant and the standard sounds, and then performed statistical analysis on the difference between the two by subtracting the standard from the deviant values for each electrode site.

Repeated-measures ANOVAs were used to evaluate behavioral and ERP results. The following factors were used: group (musicians vs. non-musicians), naturalness (NAT vs. ROT), sound type (music vs. voice), stimulus type (standards vs. deviants), hemisphere (left, right) and site. Main effects of naturalness and sound type in the N1 analysis are not discussed because these sounds categories differ in acoustic properties. However, those interactions between naturalness and other factors and between sound type and other factors that identified differences between acoustically similar sounds have been analyzed and are reported. Main effects of naturalness and sound type are reported for the P3a, P3b, and RON analyses. Significant main effects with more than two levels were evaluated with a Bonferroni post-hoc analysis. In cases where the omnibus analysis produced a significant interaction, it was further evaluated with step-down ANOVAs or t-tests, with factors specific to any given interaction. For all repeated measures with greater than one degree of freedom in the numerator, the Huynh-Feldt (H-F) adjusted *p*-values were used to determine significance (Hays, 1994). Effect sizes, indexed by the partial-eta squared statistic (η_p^2), are reported for all significant ANOVA effects. All reported t-tests are two-tailed.

3. Results

3.1 Behavioral Findings

In order to have a more objective measure of participants' musical ability, all participants were administered the Melody part of the *Music Aptitude Profile* (MAP, Gordon, 2001) test. The Melody subtest consists of pairs of short melodies – a musical question and a musical answer, according to the authors' terminology. Both melodies contain short musical phrases. In some cases, a musical answer is a melodic variation on the musical question, with extra notes added to it. In such cases, if the extra notes were removed, the question and the answer would be the same. In other cases, the musical answer is not a melodic variation on the musical question. Test takers are instructed to decide whether the musical question and the musical answer are "like" or "different." We compared the two groups on the number of incorrectly answered items out of a total of 40. Predictably, the two groups differed significantly, with overall fewer errors by musicians (mean 5.5, SD 4.33, range 0-14) compared to non-musicians (mean 10.6, SD 3.95, range 4-17), (group, $t(34)=-3.693$, $p=0.001$).

Additionally, two questions included in the musical background questionnaire were designed to probe the contribution of factors other than musical training to potential group differences. Such factors were the amount of exposure to music not directly related to training and experience with video games, with the latter having a potential to increase the speed of responses (Dye *et al.*, 2009). In order to evaluate group differences in relation to the above factors, the following questions were asked. (1) How many hours a week do you listen to music? (2) How many hours a week do you play video games? The two groups did not differ on either factor (listening to music, $t(34)=0.851$, $p=0.401$; playing video games, $t(34)=-0.515$, $p=0.61$).

A summary of ACC and RT measures for both groups is shown in Table 3. In both musicians and non-musicians deviant sounds were associated with significantly lower ACC and longer RT compared to standard sounds, thus confirming that timbre changes were in fact distracting: RT $F(1,34)=161.918$, $p<0.001$, $\eta_p^2=0.826$; ACC $F(1,34)=43.918$, $p<0.001$, $\eta_p^2=0.564$.

In regard to ACC, there was a significant effect of group, with musicians performing overall more accurately ($F(1,34)=10.661$, $p<0.01$, $\eta_p^2=0.239$). A group by sound type (voice, music) by stimulus type (standard, deviant) interaction showed a trend toward significance ($F(1,34)=3.372$, $p=0.075$, $\eta_p^2=0.09$), with musicians being equally accurate when classifying either musical or vocal deviants ($F(1,18)<1$), but with non-musicians being significantly less accurate when classifying music deviants compared to voice deviants ($F(1,16)=9.971$, $p<0.01$, $\eta_p^2=0.384$). Additionally, the naturalness (NAT, ROT) by sound type (voice, music) interaction was also significant ($F(1,34)=7.491$, $p=0.01$, $\eta_p^2=0.181$) due to the fact that in the NAT condition participants were overall more accurate when classifying vocal sounds compared to musical sounds ($F(1,36)=17.624$, $p<0.001$, $\eta_p^2=0.335$). This difference was, however, absent in the ROT condition ($F(1,36)<1$).

We also calculated a difference in accuracy between standards and deviants (see Table 3). This measure shows the degree of impairment at doing the duration discrimination task as a result of timbre change. The group difference was marginally significant ($F(1,34)=3.462$, $p=0.071$, $\eta_p^2=0.092$), with musicians' temporal discrimination accuracy being impaired to a lesser degree by the irrelevant timbre change. In addition, the group by sound type (voice, music) interaction also trended toward significance ($F(1,34)=3.372$, $p=0.075$, $\eta_p^2=0.09$). Follow-up analyses revealed that musicians were distracted to the same degree by vocal and musical timbre changes ($F(1,18)<1$), while non-musicians found musical timbre changes more distracting ($F(1,16)=7.64$, $p=0.014$, $\eta_p^2=0.323$).

In regard to RT, there was no effect of group ($F(1,34)<1$), indicating that musicians and non-musicians took on average the same amount of time to respond. Two interactions were significant. First, the sound type (voice, music) by stimulus type (standard, deviant) interaction ($F(1,34)=4.298$, $p=0.046$, $\eta_p^2=0.112$) revealed that participants responded equally fast to vocal and musical standards ($F(1,35)<1$), but were faster to respond to vocal, rather than musical, deviants ($F(1,35)=4.913$, $p=0.033$, $\eta_p^2=0.123$. Second, the naturalness (NAT, ROT) by sound type (voice, music) interaction was also significant ($F(1,34)=9.464$, $p<0.01$, $\eta_p^2=0.218$) due to faster RTs to vocal as compared to musical sounds in the NAT condition ($F(1,35)=9.395$, $p<0.01$, $\eta_p^2=0.212$).

In sum, musicians were overall more accurate at the temporal discrimination task and tended to be distracted less by irrelevant timbre change. Additionally, while musicians were equally accurate in their responses to vocal and musical deviants, non-musicians were significantly less accurate and more distracted when classifying musical as compared to vocal deviants.

3.2 ERP Findings

ERPs collected from both groups displayed the expected ERP components. In Figures 3 and 4, ERPs elicited by standards are overlaid with ERPs elicited by deviants, separately for NAT (Figure 3) and ROT (Figure 4) conditions. Figures 5 and 6 directly compare ERPs elicited in musicians and non-musicians for NAT (Figure 5) and ROT (Figure 6) sounds in order to better visualize group differences. The N1 and P3a components are marked on the Cz site, P3b – on the Pz site, and RON – on the F8 site. Below we present ERP results separately for each of the components of interest, which is followed by a summary with an emphasis on the effect of group and its interactions with other factors.

3.2.1 Neural index of early encoding of acoustic properties of sound – the N1 component

N1 peak amplitude: Musicians had a significantly larger N1 peak amplitude compared to non-musicians. This effect was present across all sites in the midline analysis ($F(1,34)=5.205$, $p=0.029$, $\eta_p^2=0.133$), over frontal, fronto-central, and central sites in the mid-lateral analysis (group by site, $F(4,136)=3.729$, $p=0.038$, $\eta_p^2=0.099$; group, $F(1,34)=4.314$-7.84, $p=0.008$-0.045, $\eta_p^2=0.113$-0.187), and over frontal and fronto-temporal sites in the lateral analysis (group by site, $F(3,102)=3.701$, $p=0.04$, $\eta_p^2=0.098$; group, $F(1,34)=3.58$-7.372, $p=0.01$-0.055, $\eta_p^2=0.104$-0.178). The effect of group did not interact with naturalness (group by naturalness: midline $F(1,34)<1$; mid-lateral, $F(1,34)<1$; lateral, $F(1,34)=1.423$, $p=0.241$). Additionally, deviants elicited a significantly larger N1 peak amplitude compared to standards (stimulus type: midline, $F(1,34)=86.22$, $p<0.001$, $\eta_p^2=0.717$; mid-lateral, $F(1,34)=130.727$, $p<0.001$, $\eta_p^2=0.794$; lateral, $F(1,34)=118.833$, $p<0.001$, $\eta_p^2=0.778$). Lastly, there were several significant results involving the effect of hemisphere over mid-lateral and lateral sites. In mid-lateral sites, the peak amplitude of N1 was overall larger over the right than over the left hemisphere sites (hemisphere, $F(1,34)=4.277$, $p=0.046$, $\eta_p^2=0.112$). In lateral sites, musical sounds (both NAT and ROT) elicited a larger N1 over the right than over the left hemisphere sites (hemisphere by sound type, $F(1,34)=7.376$, $p=0.01$, $\eta_p^2=0.178$; hemisphere, $F(1,34)=6.094$, $p=0.019$, $\eta_p^2=0.152$) while the N1 amplitude elicited by vocal sounds was similar across the two hemispheres. Lastly, the group by hemisphere by site interaction was marginally significant ($F(3,102)=3.172$, $p=0.055$, $\eta_p^2=0.085$) due to the fact that the two groups differed across a larger array of electrodes over the right as compared to the left hemisphere. Musicians had a significantly larger N1 peak amplitude than non-musicians at frontal, fronto-temporal, and temporal-parietal sites ($F(1,34)=4.294\text{-}5.953$, $p=0.02\text{-}0.046$, $\eta_p^2=0.112\text{-}0.149$) over the right hemisphere, but only at frontal sites ($F(1,34)=7.793$, $p<0.01$, $\eta_p^2=0.186$) over the left.

N1 peak latency: The two groups did not differ in the N1 peak latency. There was also no group by naturalness interaction (midline, $F(1,34)<1$; mid-lateral, $F(1,34)<1$; lateral, $F(1,34)=2.259$, $p=0.142$). The analysis yielded only one significant finding – namely, deviant sounds elicited N1 with a longer peak latency (midline, $F(1,34)=55.942$, $p<0.001$, $\eta_p^2=0.622$; mid-lateral, $F(1,34)=52.275$, $p<0.001$, $\eta_p^2=0.606$; lateral, $F(1,34)=23.724$, $p<0.001$, $\eta_p^2=0.411$).

In sum, musicians had a significantly larger N1 peak amplitude to all sound and stimulus types in both the NAT and the ROT conditions. At lateral sites, this difference was present over a larger array of electrodes over the right as compared to the left hemisphere. The two groups did not differ in the peak latency of N1.

3.2.2 Neural indices of distraction and working memory update in response to stimulus change – the P3a and P3b components—Musicians and non-musicians did not differ in the mean amplitude of the P3a component. Additionally, the factor of group did not interact with other factors. The amplitude of P3a was larger to NAT as compared to ROT sounds ($F(1,34)=25.833$, $p<0.001$, $\eta_p^2=0.432$). This difference was likely due to the reduced timbre distinctiveness between standards and deviants in the ROT condition.

The P3b analysis yielded no group effect and no interactions between group and other factors. The only significant finding was that its mean amplitude was significantly larger to NAT as compared to ROT sounds (midline, $F(1,34)=9.892$, $p<0.01$, $\eta_p^2=0.225$; mid-lateral, $F(1,34)=12.248$, $p<0.01$, $\eta_2^2=0.265$; lateral, $F(1,34)=11.458$, $p<0.01$, $\eta_p^2=0.252$). This finding is parallel to that in the P3a analysis and is likewise likely due to the reduced timbre distinctiveness between standards and deviants in the ROT condition. In sum, musicians and non-musicians did not differ in the mean amplitude of either P3a or P3b components.

3.2.3 Neural indices of recovery from distraction – the re-orienting negativity (RON)—Musicians tended to have a marginally larger mean amplitude of RON compared to non-musicians over mid-lateral and lateral sites (mid-lateral, $F(1,34)=3.211$, $p=0.082$, $\eta_p^2=0.086$; lateral, $F(1,34)=3.676$, $p=0.064$, $\eta_p^2=0.098$), suggesting a greater success at returning to the duration judgment task after the distraction took place. The peak amplitude of RON was larger to voice as compared to music deviants over midline ($F(1,34)=8.78$, $p<0.01$, $\eta_p^2=0.205$) and mid-lateral ($F(1,34)=7.508$, $p=0.01$, $\eta_p^2=0.181$) sites, with a trend in the same direction over lateral sites ($F(1,34)=3.102$, $p=0.087$, $\eta_p^2=0.084$), pointing to a greater ease at overcoming distraction when deviants were vocal as compared to musical in nature. Lastly, the mean amplitude of RON was significantly larger over the right as compared to the left hemisphere in lateral sites ($F(1,34)=21.238$, $p<0.01$, $\eta_p^2=0.384$), with a trend in the same direction in mid-lateral sites ($F(1,34)=3.683$, $p=0.063$, $\eta_p^2=0.098$).

3.3 Correlations

In order to determine whether an enhanced N1 is correlated with behavioral measures of musical expertise, we examined a connection between the N1 peak amplitude and the following measures: onset of musical training, years of musical training, MAP scores, self-rated musical proficiency, and the number of hours listening to music per week. The N1 peak amplitude averages were calculated for midline, mid-lateral, and lateral sites for each participant. Separate regression analyses were performed between each of the above behavioral measures and the N1 average for each scalp area. Because the amplitude of N1 was significantly smaller in response to standards compared to deviants (likely due to the refractoriness of the neurons responding to repeating standard sounds (Näätänen & Picton, 1987)), we conducted separate regression analyses on N1 to standards and on N1 to deviants. All reported *p* values are two-tailed.

N1 to deviants—In the NAT condition, individuals with higher self-rated musical proficiency had a significantly larger N1 peak amplitude to both music and voice deviants over the mid-lateral (music deviants, *r*=0.371, *p*=0.026; voice deviants, *r*=0.338, *p*=0.044) and midline (music deviants, *r*=0.351, *p*=0.036; voice deviants, *r*=0.342, *p*=0.041) sites, with a trend in the same direction over the lateral sites (music deviants, *r*=0.315, *p*=0.061; voice deviants, *r*=0.281, *p*=0.097). Additionally, the N1 elicited by music deviants was larger over the lateral sites (*r*=0.357, *p*=0.032) in individuals with higher MAP scores. A relationship between the N1 peak amplitude to voice deviants and MAP scores showed a similar trend (*r*=0.291, *p*=0.085). None of the results in the ROT condition reached significance.

N1 to standards—In the NAT condition, individuals with higher self-rated musical proficiency had a significantly larger N1 peak amplitude to both music and voice standards over the midline (music standards, *r*=0.335, *p*=0.046; voice standards, *r*=0.402, *p*=0.015) and mid-lateral (music standards, *r*=0.331, *p*=0.049; voice standards *r*=0.385, *p*=0.02) sites. Individuals with higher MAP scores had a larger N1 to voice standards (*r*=0.342, *p*=0.041) and a marginally larger N1 to music standards *(r*=0.295, *p*=0.081) over the lateral sites. In the ROT condition, only one relationship reached significance; namely, individuals with higher self-rated musical proficiency had larger N1 to ROT music standards over the midline (*r*=0.361, *p*=0.03) and mid-lateral (*r*=0.331, *p*=0.049) sites.

3.4 Results summary

The above sections have listed all significant results of the study. Here we summarize them again, with the focus on the findings that bear directly on the main questions of the study and that will be further evaluated in the Discussion. These findings are as follows. Behavioral measures revealed that all participants were faster and more accurate when classifying vocal as compared to musical sounds, both standards and deviants. Musicians were overall more accurate when making sound duration judgments. They responded equally accurately to vocal and musical deviants, while non-musicians were less accurate and more delayed in their responses to music as compared to voice deviants. Electrophysiological measures showed a significantly larger N1 peak amplitude in musicians, regardless of the nature of the stimulus (standard vs. deviant, voice vs. music, natural vs. spectrally-rotated). This group difference was present across a larger number of electrodes over the right as compared to the left hemisphere. The N1 peak amplitude to NAT sounds was positively correlated with self-rated music proficiency and performance on the MAP test. The two groups did not differ in the mean amplitude of the P3a and the P3b components. However, musicians showed a marginally larger RON. The mean amplitude of RON was significantly greater over the right hemisphere.

4. Discussion

4.1 Do musicians show an enhanced encoding of the human voice?

We asked whether early sensory encoding of vocal and completely novel sounds may be enhanced in amateur musicians compared to non-musicians (e.g., Pantev *et al.*, 1998; Shahin *et al.*, 2003; Shahin *et al.*, 2004; Fujioka *et al.*, 2006). We compared the N1 peak amplitude and peak latency elicited by musical and vocal sounds, as well as by their spectrally rotated versions, as a measure of such sensory encoding. We found that musicians had a significantly larger N1 peak amplitude. This effect did not interact either with sound type (voice, music) or with naturalness (NAT, ROT). Instead, it was present across the board, even in response to completely novel and never before heard spectrally-rotated sounds.

The lack of timbre-specificity in our results suggests that the enhancement in the N1 component shown by musicians is not due to the perceptual similarity between musical and vocal timbres; instead, it is likely that musical training leads to a more general enhancement in the encoding of at least some acoustic features that are shared by perceptually dissimilar but acoustically complex sound categories. One of the acoustic features whose perception may be fine-tuned by musical training is spectral complexity. For example, Shahin and colleagues manipulated the number of harmonics in a piano note and reported a larger P2m to tones with a higher number of harmonics in trained pianists (Shahin *et al.*, 2005). Although in Shahin and colleagues' study the spectral complexity was manipulated in a musical note only, it is likely that musicians show a similar advantage in other types of sounds. Previous studies consistently identified several spectrum-related acoustic features that contribute to the perception of timbre, such as the spectral center of gravity and spectrum fine structure (Caclin *et al.*, 2005; Caclin *et al.*, 2008). Therefore, greater sensitivity to such spectral properties of sound may lead to better neural encoding and enhanced perceptual processing of various timbres, both familiar and novel. This issue requires further study.

Another interpretation of a larger N1 peak amplitude in musicians is possible – namely, it may index a greater attentional allocation to auditory stimuli in this group of participants. Whether the enhancement of early sensory components in musicians is due at least in part to differences in attentional modulation is a topic of ongoing debate. However, in a recent study, Baumann and colleagues (Baumann *et al.*, 2008) directly compared the N1 and P2 components in musicians to the same sounds in two different tasks. In one, participants had to attend to certain sound properties (such as pitch and timbre) while in another they did not. The authors demonstrated that intentionally directing attention to sound properties did not increase the amplitude of the N1 and P2 components and, therefore, concluded that the previously reported enhancement of these components in musicians is due to greater auditory expertise and not to differences in attentional allocation between musicians and non-musicians.

Studies on vocal and musical timbre perception tend to focus either on musical timbre perception in musicians or on vocal timbre perception in the general population; however, few bridge these two broad areas. For example, Pantev and colleagues (Pantev *et al.*, 2001) compared magnetoencephalographic (MEG) recordings to violin and trumpet notes in violinists and trumpeters and found that the amplitude of N1m was larger to violin notes than to trumpet notes in the group of violinists and larger to trumpet notes than to violin notes in the group of trumpeters. The authors concluded that their results support timbre-specific enhancement of brain responses in musicians, which was dependent on the instrument of training. This finding has been supported by other studies. Shahin and colleagues (Shahin *et al.*, 2008) reported greater induced gamma band activity in pianists and violinists, specifically for the instrument of practice. Musicians also show greater activation in an extensive network of brain regions in the left hemisphere (including the precentral gyrus, the inferior frontal gyrus, the inferior parietal lobule, and the medial frontal gyrus) when listening to a musical piece played in their instrument of training as compared to a different instrument (Margulis *et al.*, 2009). Such sensitivity to the timbre of the instrument of training is already evident at the subcortical level as has been shown by Strait and colleagues who reported that pianists had a greater degree of correspondence between brainstem responses and the acoustic wave of piano notes compared to similar correspondences for bassoon or tuba notes (Strait *et al.*, 2012).

On the other hand, a small but growing number of studies have focused on the timing and specificity of voice-elicited ERPs. First studies on the electrophysiological signature of voice perception reported the presence of the Voice-Sensitive Response (VSR) peaking at approximately 320 ms post-stimulus onset (Levy *et al.*, 2001; 2003) and thought to reflect the allocation of attention to voice stimuli. Levy and colleagues were also among the first to directly compare ERP responses to vocal and musical sounds in non-musicians and to demonstrate that such responses were overall quite similar, especially when participants did not attend to stimuli or did not focus on timbre during stimuli processing. More recent work suggests that voice-specific auditory processing happens significantly earlier than VSR, approximately in the time range of the P2 ERP component (e.g., Charest *et al.*, 2009; Rogier *et al.*, 2010; Capilla *et al.*, in press), although the timing of this “fronto-temporal positivity to voice” (FTPV) varies somewhat from study to study. Further support for the relatively early processing of vocal properties comes from studies reporting that gender and voice identity are detected at approximately the same time with the occurrence of FTPV (e.g., Zäske *et al.*, 2009; Schweinberger *et al.*, 2011; Latinus & Taylor, 2012).

To the best of our knowledge, to date, just one study has examined the effect of musical training on voice perception (Chartrand & Belin, 2006). It found that musicians were more accurate than non-musicians in discriminating vocal and musical timbres, but took longer to respond. The results of our study begin to describe the neural processes potentially underlying such advantage in musicians and contribute to previous research by bridging the two literatures discussed above. Our findings do not contradict earlier reports of timbre-specific enhancement in musicians but extend them in an important way. By including vocal and highly novel timbres in our experimental design, we were able to examine the degree to which the enhancement of early sound encoding due to musical training may generalize to other complex sound categories. The fact that musicians displayed an enhanced N1 to spectrally-rotated sounds and that the two groups differed during a rather early time window (in the 150-220 ms post-stimulus onset range) strongly suggests that musical training is associated not only with timbre-specific enhancement of neural responses as described in earlier studies, but also with a more general enhancement in the encoding of acoustic properties of sounds, even when such sounds are perceptually dissimilar to the instrument(s) of training. Such enhanced encoding of acoustic properties of sound during early neural processing (in the N1 time range) might then contribute to better perceptual processing of various timbres during a slightly later processing window (the P2 time range), as reported in previous studies.

The finding of a larger amplitude of the N1 component over the right as compared to the left hemisphere sites and of a more widespread group difference in the N1 peak amplitude over the right hemisphere in our study is noteworthy. Lateralization effects in ERP results should be interpreted with caution; however, our results do agree with reports of greater right hemisphere involvement in the processing of spectral information and of timbre in particular (e.g., Belin *et al.*, 2000; Zatorre & Belin, 2001; von Kriegstein *et al.*, 2003).

While the N1 enhancement in musicians was present to all sound types, the relationship between its peak amplitude and measures of musical proficiency was limited to the NAT condition. More specifically, individuals who rated their own musical ability more highly had a larger N1 peak amplitude to both music and voice deviants. Additionally, individuals with higher MAP scores had higher N1 peak amplitude to music deviants. A similar, but a weaker relationship was also present between MAP scores and N1 to voice deviants. A relationship between N1 and either the age at onset of training or the duration of training was not significant. In part this may be due to the fact that we tested amateur musicians, who on average started their training later than what would be typical for professional musicians. Overall, however, reports of correlation between either the age at the onset of musical training or the duration of such training and the enhancement of early ERP responses are not consistent (e.g., Pantev *et al.*, 1998; Shahin *et al.*, 2003; Musacchia *et al.*, 2007).

Our evaluation of timbre encoding in musicians and non-musicians has its limitations. Our main task probed the ability of the two groups of participants to resist distraction and did not measure overt timbre perception. Therefore, whether enhanced N1 peak amplitude to complex sounds in musicians actually translates into better timbre identification and/or discrimination requires future studies. Related to the above point is the fact that the design of our study required that we use only a small set of sounds to represent vocal and musical timbres. In contrast, studies of the FTPV component used a large range of vocal and non-vocal sounds. Future studies that use a larger set of timbre examples and focus on the FTPV component may help determine whether musicians' neural encoding of voices as a perceptual category (compared to voices' acoustic properties as in the current study) is superior to that in non-musicians.

In sum, musicians showed an enhanced N1 ERP component not only to musical and vocal sounds but also to never before heard spectrally-rotated sounds. This finding suggests that musical training is associated with a general enhancement in the neural encoding of acoustic properties of complex sounds, which are not tightly coupled with the specific timbres to which musicians are exposed during training.

4.2 Do musicians show an enhanced inhibition of irrelevant auditory change?

Previous research has demonstrated that musical training may sharpen not only one's perceptual skills but also one's ability to allocate and sustain attention (Pallesen *et al.*, 2010; Moreno *et al.*, 2011; Strait & Kraus, 2011). We asked whether musical training may also enhance one's ability to resist distraction by task-irrelevant auditory change. To do so, we used an auditory distraction paradigm developed by Schröger and colleagues (Schröger & Wolff, 1998; 2000), in which participants were asked to classify sounds as either short or long and ignore a rare and task-irrelevant change in sounds' timbre. Both groups were able to do the duration discrimination task successfully; however, musicians performed overall better than non-musicians. Given the important role that sound duration plays in music, this finding is not surprising and is in agreement with earlier reports (Güçlü *et al.*, 2011). Although the overall group difference in the degree of distraction by all types of deviants fell just short of the significance cut-off, in general, musicians' accuracy tended to be affected less by irrelevant timbre change. Further, musicians were equally accurate at classifying vocal and musical deviants according to the sound length, and were distracted to the same degree by the two types of deviants. Non-musicians, on the other hand, found musical deviant classification more challenging and were distracted by musical deviants more than by vocal deviants. These findings suggest that while musical training may potentially enhanced one's ability to resist auditory distraction in general, this skill appears to depend on the familiarity with the irrelevant sound dimension along which distracting changes occur. Thus, musicians clearly outperformed non-musicians when deviants were musical sounds, but the two groups performed similarly when deviants were voices.

We also examined three ERP measures associated with distraction – namely, the P3a, P3b, and RON components. The P3a and P3b components did not differentiate the two groups, suggesting that the processes of deviance detection and working memory update in response to auditory change were similar in musicians and non-musicians. However, the RON component, thought to index the successful return to the task at hand after distraction took place, was marginally larger in musicians compared to non-musicians, suggesting that overall musicians tended to be more successful at returning to the duration discrimination task after being distracted by the irrelevant timbre change. This finding agrees with the accuracy data described above.

The amplitude of the RON component was significantly larger over the right hemisphere across all analyses – a finding, which, to the best of our knowledge, has not been reported in earlier studies of RON. The difference likely lies in the nature of our stimuli since most previous studies used simple tones of different frequencies. A greater engagement of the right hemisphere in our study might suggest that those aspects of the executive function that are involved in the recovery from distraction are not independent of the processes underlying timbre perception. On this account, the greater amplitude of RON over the right hemisphere may reflect increased inhibitory activation of the right hemisphere brain regions previously implicated in the processing of spectral complexity and timbre as participants disengage their attention from sound timbre and re-focus it on sound duration. This question requires further study.

Lastly, RON was larger to vocal as compared to musical deviants, lending support to the behavioral finding that voice deviants were overall less distracting than music deviants. One reason for the greater ease of screening out vocal changes may be the fact that regardless of our musical background we all are voice experts (e.g., Chartrand *et al.*, 2008; Latinus & Belin, 2011). Indeed, we encounter the need to both identify the talker and ignore talker variability in speech on a daily basis and thus have extensive experience in separating talker-related information from the rest of the speech signal. Recent neuroimaging and neuropsychological studies suggest that different aspects of voice perception (those related to speech, affect, and talker recognition) may in fact be processed in semi-independent neural structures (e.g., Belin *et al.*, 2000; von Kriegstein *et al.*, 2003; Belin *et al.*, 2004; Garrido *et al.*, 2009; Spreckelmeyer *et al.*, 2009; Hailstone *et al.*, 2010; Gainotti, 2011). Furthermore, sensitivity to voice information develops exceptionally early. For example, the ability to discriminate between the voice of one’s mother and the voice of a stranger emerges before birth (Ockleford *et al.*, 1988; Kisilevsky *et al.*, 2003). By 4 to 5 months of age, infants begin to show the fronto-temporal positivity to voice (Rogier *et al.*, 2010) and by seven months of age demonstrate a greater right-hemisphere brain activity in response to voice as compared to other sounds, similar to that found in adults (Grossmann *et al.*, 2010). Finally, by one year of age infants are able to follow others’ voice direction (Rossano *et al.*, in press), suggesting that they are capable of using voice information alone for establishing joint attention. Such expertise at voice processing might have rendered the task of separating vocal information from sound duration in our experiment relatively easy for both groups. However, only musicians had had extensive experience in extracting sound duration from different musical timbres prior to participating in the study, which has likely contributed to their better ability to identify sound duration of musical notes even when the latter were distracting deviants.

In sum, analysis of behavioral and electrophysiological measures indicates that musicians’ accuracy tended to suffer less from the change in sounds’ timbre, especially when deviants were musical notes. This behavioral finding was accompanied by a larger re-orienting negativity in musicians, suggesting that their advantage lies in a more efficient disengagement of attention from the distracting auditory dimension. Our findings suggest that one's ability to recover from distraction depends at least in part on the extent of prior experience with the auditory dimension of change.

5. Conclusion

Musicians exhibited a larger N1 ERP component not only to musical and vocal sounds, but also to never before heard spectrally-rotated sounds. This finding suggests that musical training is associated with a general enhancement in the early neural encoding of complex sounds, even when such sounds' timbre is dissimilar to the timbre of the instrument of training. While the N1 enhancement in musicians was present across the board, their ability to ignore irrelevant auditory change surpassed that of non-musicians only when distractors were music sounds, pointing to the role of familiarity with a specific timbre in this skill.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This project was supported in part by award number P30DC010745 from the National Institute on Deafness and Other Communicative Disorders. The content is solely the responsibility of the authors and does not necessarily represent the official view of the National Institute on Deafness and Other Communicative Disorders or the National Institutes of Health. We are grateful to Dan Noland for his invaluable help with programming and to Jayaganesh Swaminathan for creating spectrally-rotated sounds. The authors report no conflict of interest.

References

Human Experimentation: Code of Ethics of the World Medical Association. British Medical Journal. 1964; 2:177. [PubMed: 14150898]

American Electroencephalographic Society. Guideline thirteen: Guidelines for standard electrode placement nomenclature. Journal of Clinical Neurophysiology. 1994; 11:111–113. [PubMed: 8195414]

Baumann S, Meyer M, Jäncke L. Enhancement of auditory-evoked potentials in musicians reflects an influence of exertise but not selective attention. Journal of Cognitive Neuroscience. 2008; 20:2238–2249. [PubMed: 18457513]

Belin P, Fecteau S, Bédard C. Thinking the voice: neural correlates of voice perception. Trends in Cognitive Sciences. 2004; 8:129–135. [PubMed: 15301753]

Belin P, Zatorre RJ, Fafaille P, Ahad P, Pike B. Voice-selective areas in human auditory cortex. Nature. 2000; 403:309–312. [PubMed: 10659849]

Blesser B. Speech perception under conditions of spectral transformation: I. Phonetic characteristics. Journal of Speech and Hearing Research. 1972; 15:5–41. [PubMed: 5012812]

Boersma, P.; Weenink, D. Praat: doing phonetics by computer (version 5.3) [Computer program]. 2011. Retrieved from http://www.praat.org

Caclin A, McAdams S, Smith BK, Giard M-H. Interactive processing of timbre dimensions: An exploration with event-related potentials. Journal of Cognitive Neuroscience. 2008; 20:1–16. [PubMed: 17919082]

Caclin A, McAdams S, Smith BK, Winsberg S. Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. Journal of the Acoustical Society of America. 2005; 118:471–482. [PubMed: 16119366]

Capilla A, Belin P, Gross J. The early spatio-temporal correlates and task independence of cerebral voice processing studied with MEG. Cerebral Cortex. (in press).

Charest I, Pernet CR, Rousselet GA, Quiñones I, Latinus M, Fillion-Bilodeau S, Chartrand J-P, Belin P. Electrophysiological evidence for an early processing of human voices *BMC Neuroscience*. BioMed Central. 2009

Chartrand J-P, Belin P. Superior voice timbre processing in musicians. Neuroscience Letters. 2006; 405:164–167. [PubMed: 16860471]

Chartrand J-P, Peretz I, Belin P. Auditory recognition expertise and domain specificity. Brain Research. 2008; 1220:191–198. [PubMed: 18299121]

Donchin E. Surprise... Surprise? Psychophysiology. 1981; 18:493–513. [PubMed: 7280146]

Donchin E, Coles MGH. Precommentary: Is the P300 component a manifestation of context updating? Behavioral and Brain Sciences. 1988; 11:493–513.

Dye MWG, Green CS, Bavelier D. Increasing speed of processing with action video games. Current Directions in Psychological Science. 2009; 18:321–326. [PubMed: 20485453]

Fujioka T, Ross B, Kakigi R, Pantev C, Trainor LJ. One year of musical training affects development of audiory cortical-evoked fields in young children. Brain. 2006; 129:2593–2608. [PubMed: 16959812]

Gainotti G. What the study of voice recognition in normal subjects and brain-damaged patients tells us about models of familiar people recognition. Neuropsychologia. 2011; 49:2273–2282. [PubMed: 21569784]

Garrido L, Eisner F, McGettigan C, Stewart L, Sauter D, Hanley JR, Schweinberger SR, Warren JD, Duchaine B. Developmental phonagnosia: A selective deficit of vocal identity recognition. Neuropsychologia. 2009; 47:123–131. [PubMed: 18765243]

Gordon, EE. Music Aptitude Profile. GIA Publications, Inc.; Chicago, IL: 2001.

Grossmann T, Oberecker R, Koch SP, Friederici AD. The developmental origins of voice processing in the human brain. Neuron. 2010; 65:852–858. [PubMed: 20346760]

Güçlü B, Sevinc E, Canbeyli R. Duration discrimination by musicians and nonmusicians. Psychological Reports. 2011; 108:675–687. [PubMed: 21879613]

Hailstone JC, Crutch SJ, Vestergaard MD, Patterson RD, Warren JD. Progressive associative phonagnosia: A neurpsychological analysis. Neuropsychologia. 2010; 48:1104–1114. [PubMed: 20006628]

Halpern AR, Zatorre RJ, Bouffard M, Johnson JA. Behavioral and neural correlates of perceived and imagined musical timbre. Neuropsychologia. 2004; 42:1281–1292. [PubMed: 15178179]

Handel, S. Listening: an introduction to the perception of auditory events. The MIT Press; Cambridge: 1989.

Hays, M. Statistics. Harcourt Brace College Publishers; Fort Worth, TX: 1994.

Horváth J, Sussman ES, Winkler I, Schröger E. Preventing distraction: Assessing stimulus-specific and general effects of the predictive cueing of deviant auditory events. Biological Psychology. 2011; 87:35–48. [PubMed: 21310210]

Horváth J, Winkler I, Bendixen A. Do N2/MMN, P3a, and RON form a strongly coupled chain reflecting the three stages of auditory distraction? Biological Psychology. 2008; 79:139–147. [PubMed: 18468765]

Kisilevsky BS, Hains SMJ, Lee K, Xie X, Huang H, Ye HH, Zhang K, Wang Z. Effects of experience on fetal voice recognition. Psychological Science. 2003; 14:220–224. [PubMed: 12741744]

Kreiman, J. Listening to voices: Theory and practice in voice perception research. In: Johnson, K.; Mullennix, JW., editors. Talker Variability in Speech Processing. Academic Press; New York: 1997.

Latinus M, Belin P. Human voice perception. Current Biology. 2011; 21:R143–R145. [PubMed: 21334289]

Latinus M, Taylor MJ. Discriminating male and female voices: differentiating pitch and gender. Brain Topography. 2012; 25:194–204. [PubMed: 22080221]

Levy DA, Granot R, Bentin S. Processing specificity for human voice stimuli: electrophysiological evidence. NeuroReport. 2001; 12:2653–2657. [PubMed: 11522942]

Levy DA, Granot R, Bentin S. Neural sensitivity to human voices: ERP evidence of task and attentional influences. Psychophysiology. 2003; 40:291–305. [PubMed: 12820870]

Luck, S. An Introduction to the Event-Related Potential Technique. The MIT Press; Cambridge, MA: 2005.

Mager R, Falkenstein M, Störmer R, Brand S, Müller-Spahn, Bullinger AH. Auditory distraction in young and middle-aged adults: A behavioral and event-related potential study. Journal of Neural Transmission. 2005; 112:1165–1176. [PubMed: 15614427]

Margulis EH, Mlsna LM, Uppunda AK, Parrish TB, Wong PCM. Selective neurophysiologic reponses to music in instrumentalists with different listening biographies. Human Brain Mapping. 2009; 30:267–275. [PubMed: 18072277]

McAdams S, Winsberg S, Donnadieu S, De Soete G, Krimphoff J. Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. Psychological Research. 1995; 58:177–192. [PubMed: 8570786]

Moreno S, Bialystok E, Barac R, Schellenberg EG, Cepeda NJ, Chau T. Short-tem music training enhances verbal intelligence and executive function. Psychological Science. 2011:1425–1433. [PubMed: 21969312]

Musacchia G, Sams M, Skoe E, Kraus N. Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. Proceedings of the National Academy of Sciences. 2007; 104:15894–15898.

Näätänen R. The Mismatch Negativity: A powerful tool for cognitive neuroscience. Ear and Hearing. 1995; 16:6–18. [PubMed: 7774770]

Näätänen R. The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). Psychophysiology. 2001; 38:1–21. [PubMed: 11321610]

Näätänen R, Alho K. Mismatch negativity - a unique measure of sensory processing in audition. International Journal of Neuroscience. 1995; 80:317–337. [PubMed: 7775056]

Näätänen R, Picton T. The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. Psychophysiology. 1987; 24:375–425. [PubMed: 3615753]

Ockleford EM, Vince MA, Layton C, Reader MR. Responses of neonates to parents' and others' voices. Early Human Development. 1988; 18:27–36. [PubMed: 3234282]

Pallesen KJ, Brattico E, Bailey CJ, Korvenoja A, Koivisto J, Gjedde A, Carlson S. Cognitive control in auditory working memory is enhanced in musicians. PLoS One. 2010; 5:e11120. [PubMed: 20559545]

Pantev C, Oostenveld R, Engelien A, Ross B, Roberts LE, Hoke M. Increased auditory cortical representation in musicians. Nature. 1998; 392:811–813. [PubMed: 9572139]

Pantev C, Roberts LE, Schultz M, Engelien A, Ross B. Timbre-specific enhancement of auditory cortical representations in musicians. NeuroReport. 2001; 12:169–174. [PubMed: 11201080]

Peretz I, Belleville S, Fontaine S. Dissociantions entre musique et langage apres atteinte cerebrale: un nouveau cas d'amusie sans aphasie. Canadian journal of experimental psychology. 1997; 51:354. [PubMed: 9687196]

Peretz I, Kolinsky R, Tramo M, Labrecque R, Hublet C, Demeurisse G, Belleville S. Functional dissociations following bilateral lesions of auditory cortex. Brain. 1994; 117:1283–1301. [PubMed: 7820566]

Picton T, Bentin S, Berg P, Donchin E, Hillyard SA, Johnson R Jr. Miller GA, Ritter W, Ruchkin DS, Rugg MD, Taylor MJ. Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. Psychophysiology. 2000; 37:127–152. [PubMed: 10731765]

Picton TW. The P300 wave of the human event-related potential. Journal of Clinical Neurophysiology. 1992; 9:456–479. [PubMed: 1464675]

Polich J. Updating P300: An integrative theory of P3a and P3b. Clincial Neurophysiology. 2007; 118:2128–2148.

Pollich, J. Theoretical overview of P3a and P3b. In: Polich, J., editor. Detection of Change: Event-Related Potential and fMRI Findings. Kluwer Academic Publishers; NewYork: 2003.

Rogier O, Roux S, Belin P, Bonnet-Brilhault F, Bruneau N. An electrophysiological correlate of voice processing in 4- to 5-year old children. International Journal of Psychophysiology. 2010; 75:44–47. [PubMed: 19896509]

Rossano F, Carpenter M, Tomasello M. One-year-old infants follow others’ voice direction. Psychological Science. (in press).

Samson S, Zatorre RJ. Contribution of the right temporal lobe to musical timbre discrimination. Neuropsychologia. 1994; 32:231–240. [PubMed: 8190246]

Samson S, Zatorre RJ, Ramsay JO. Deficits of musical timbre perception after unilateral temporal-lobe lesion revealed with multidimensional scaling. Brain. 2002; 125:511–523. [PubMed: 11872609]

Schröger E, Wolff C. Behavioral and electrophysiological effects of task-irrelevant sound change: A new distraction paradigm. Cognitive Brain Research. 1998; 7:71–87. [PubMed: 9714745]

Schröger E, Wolff C. Auditory distraction: Event-related potential and behavioral indices. Clinical Neurophysiology. 2000; 111:1450–1460. [PubMed: 10904227]

Schweinberger SR, Walther C, Zäske R. Neural correlates of adaptation to voice identity. British Journal of Psychology. 2011; 102:748–764. [PubMed: 21988382]

Shahin AJ, Bosnyak DJ, Trainor LJ, Roberts LE. Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. The Journal of Neuroscience. 2003; 23:5545–5552. [PubMed: 12843255]

Shahin AJ, Roberts LE, Chau W, Trainor LJ, Miller LM. Music training leads to the development of timbre-specific gamma band activity. NeuroImage. 2008; 41:113–122. [PubMed: 18375147]

Shahin AJ, Roberts LE, Pantev C, Trainor LJ, Ross B. Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. NeuroReport. 2005; 16

Shahin AJ, Roberts LE, Trainor LJ. Enhancement of auditory cortical development by musical experience in children. NeuroReport. 2004; 15:1917–1921. [PubMed: 15305137]

Spreckelmeyer KN, Kutas M, Urbach T, Altenmüller E, Münte TF. Neural processing of vocal emotion and identity. Brain and Cognition. 2009; 69:121–126. [PubMed: 18644670]

Strait DL, Chan D, Ashley R, Kraus N. Specialization among the speicalized: auditory brainstem function is tuned in to timbre. Cortex. 2012; 48:360–362. [PubMed: 21536264]

Strait DL, Kraus N. Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. Frontiers in Psychology. 2011; 2:1–10. [PubMed: 21713130]

Sussman ES. A new view on the MMN and attention debate: The role of context in processing auditory events. Journal of Psychophysiology. 2007; 21:164–175.

Trainor LJ, Shahin AJ, Roberts LE. Understanding the benefits of musical training: Effects on oscillatory brain activity. Annals of the New York Academy of Sciences. 2009; 1169:133–142. [PubMed: 19673769]

von Kriegstein K, Eger E, Kleinschmidt A, Giraud AL. Modulaton of neural responses to speech by directing attention to voices or verbal content. Cognitive Brain Research. 2003; 17:48–55. [PubMed: 12763191]

Wronka E, Kaiser J, Coenen AML. Neural generators of the auditory evoked potential components P3a and P3b. Acta Neurobiologiae Experimentalis. 2012; 72:51–64. [PubMed: 22508084]

Zäske R, Schweinberger SR, Kaufmann J, Kawahara H. In the ear of the beholder: neural correlates of adaptation to voice gender. European Journal of Neuroscience. 2009; 30:527–534. [PubMed: 19656175]

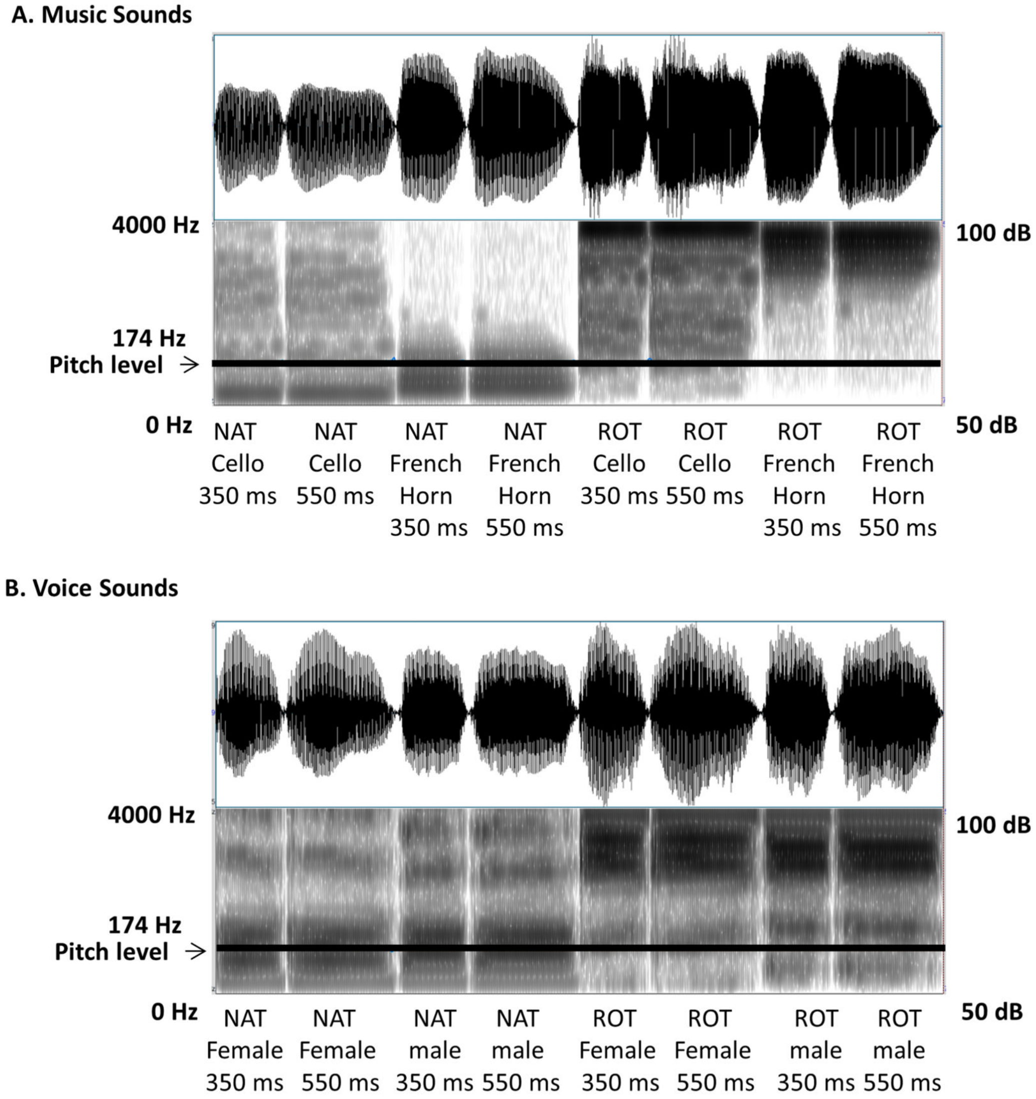

Figure 1.
Waveforms and spectrograms of stimuli

Waveforms and spectrograms of both music (Panel A) and voice (Panel B) stimuli are shown. Pitch level is marked with a horizontal line. Pitch was kept constant at 174 Hz for all sounds. Frequency range is specified on the left and intensity range - on the right. Note a similarity in acoustic complexity and temporal envelope between NAT and ROT sounds.

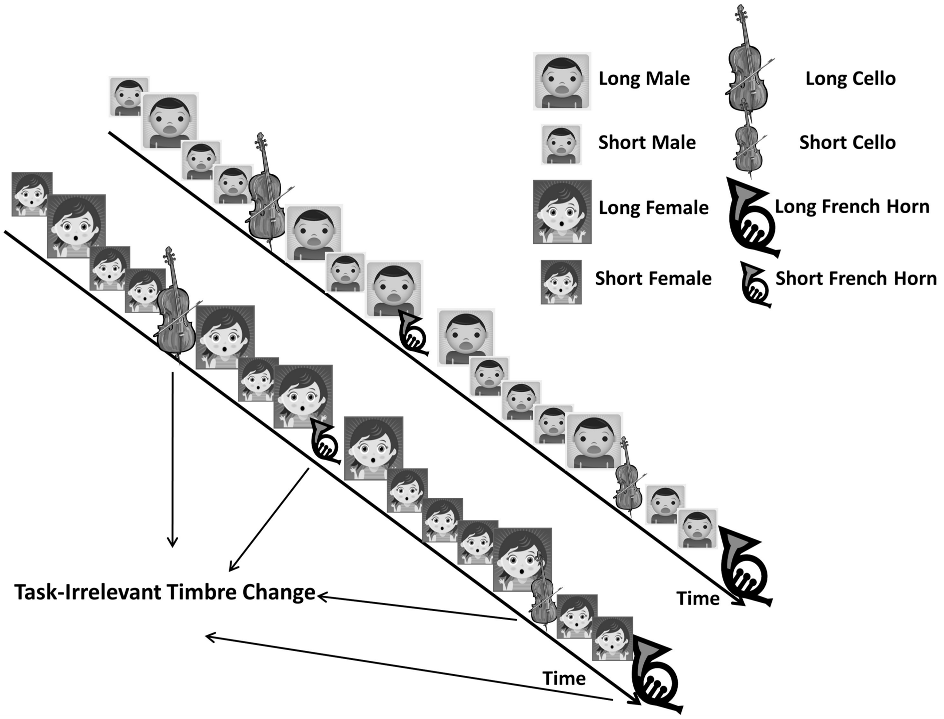

Figure 2.

Experimental design

The two timelines exemplify blocks with frequent voice sounds and rare music deviants. Large images represent long sounds while small images represent short sounds. As can be seen, both frequent and rare sounds were of two lengths. Individuals pressed one button for short sounds and another button for long sounds. A change in the sound timbre – from voice to music in this case – was irrelevant to the duration discrimination task. Blocks with frequent music sounds and rare voice deviants had an identical design.

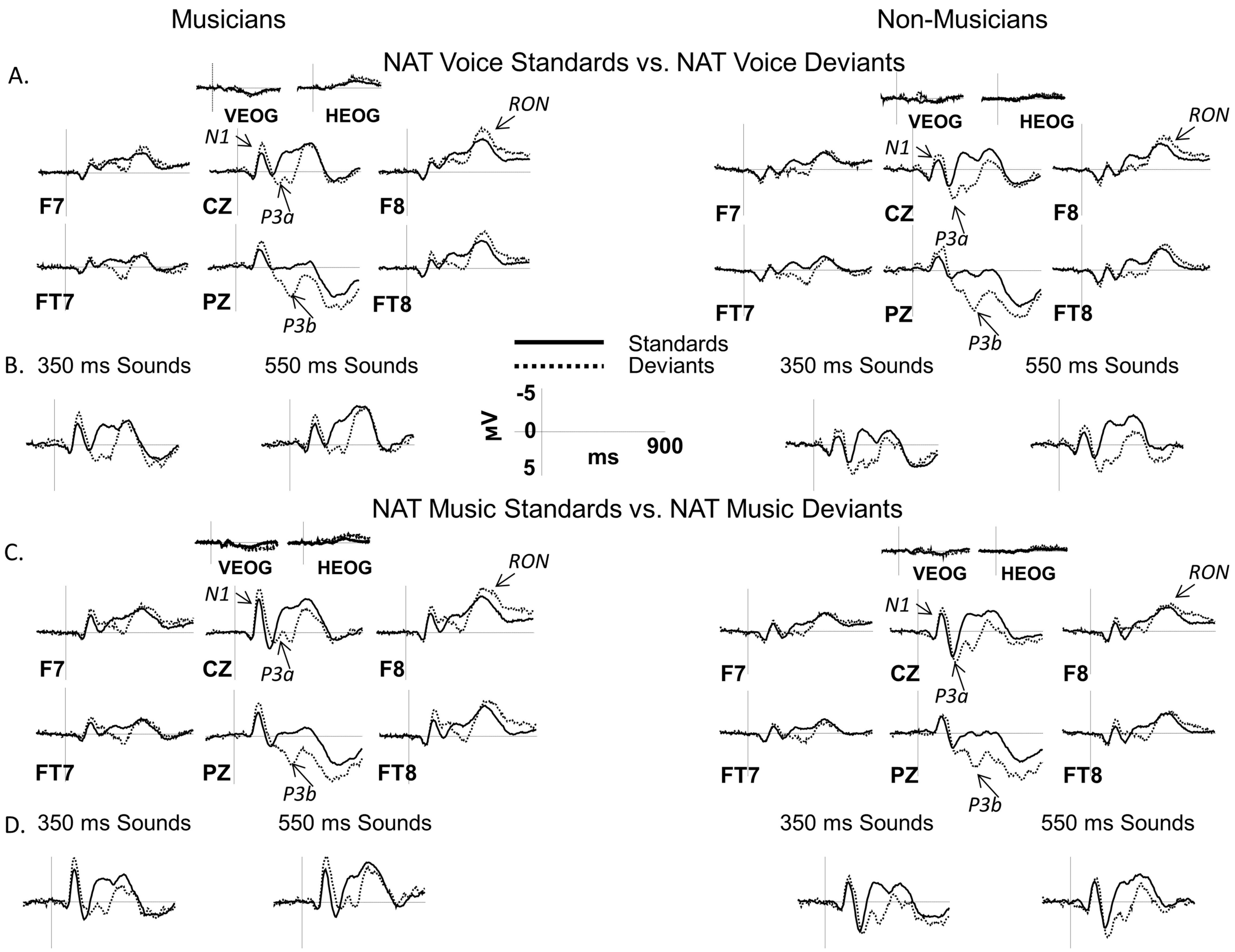

Figure 3.

ERP results in the NAT condition: standards vs. deviants

ERPs elicited by vocal and musical sounds in the NAT condition are shown over a representative set of electrodes. A, C: ERPs elicited by deviants are overlaid with those elicited by standards separately for musicians (left side) and non-musicians (right side) and for vocal (panel A) and musical (panel C) sounds. ERPs were averaged across short and long sound durations for all comparisons. B, D: Separate ERP grand averages for short (350 ms) and long (550 ms) sounds are shown at the CZ site for vocal (panel B) and musical (panel D) sounds. Analyzed ERP components are marked on multiple electrode sites. Negative potentials are plotted upward. Time 0 ms indicates the onset of the stimulus.

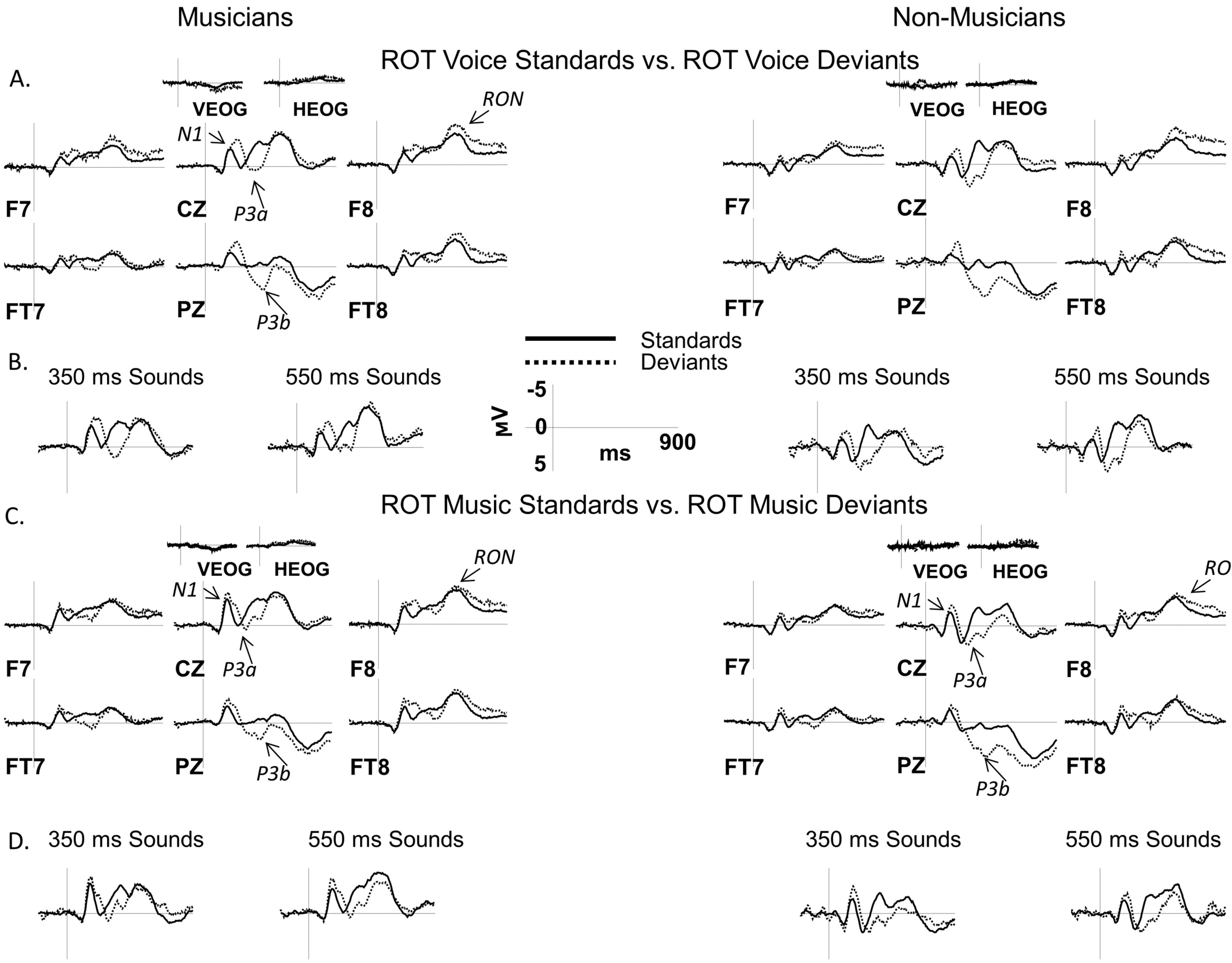

Figure 4.

ERP results in the ROT condition: standards vs. deviants

ERPs elicited by vocal and musical sounds in the ROT condition are shown over a representative set of electrodes. A, C: ERPs elicited by deviants are overlaid with those elicited by standards separately for musicians (left side) and non-musicians (right side) and for vocal (panel A) and musical (panel C) sounds. ERPs were averaged across short and long sound durations for all comparisons. B, D: Separate ERP grand averages for short (350 ms) and long (550 ms) sounds are shown at the CZ site for vocal (panel B) and musical (panel D) sounds. Analyzed ERP components are marked on multiple electrode sites. Negative potentials are plotted upward. Time 0 ms indicates the onset of the stimulus.

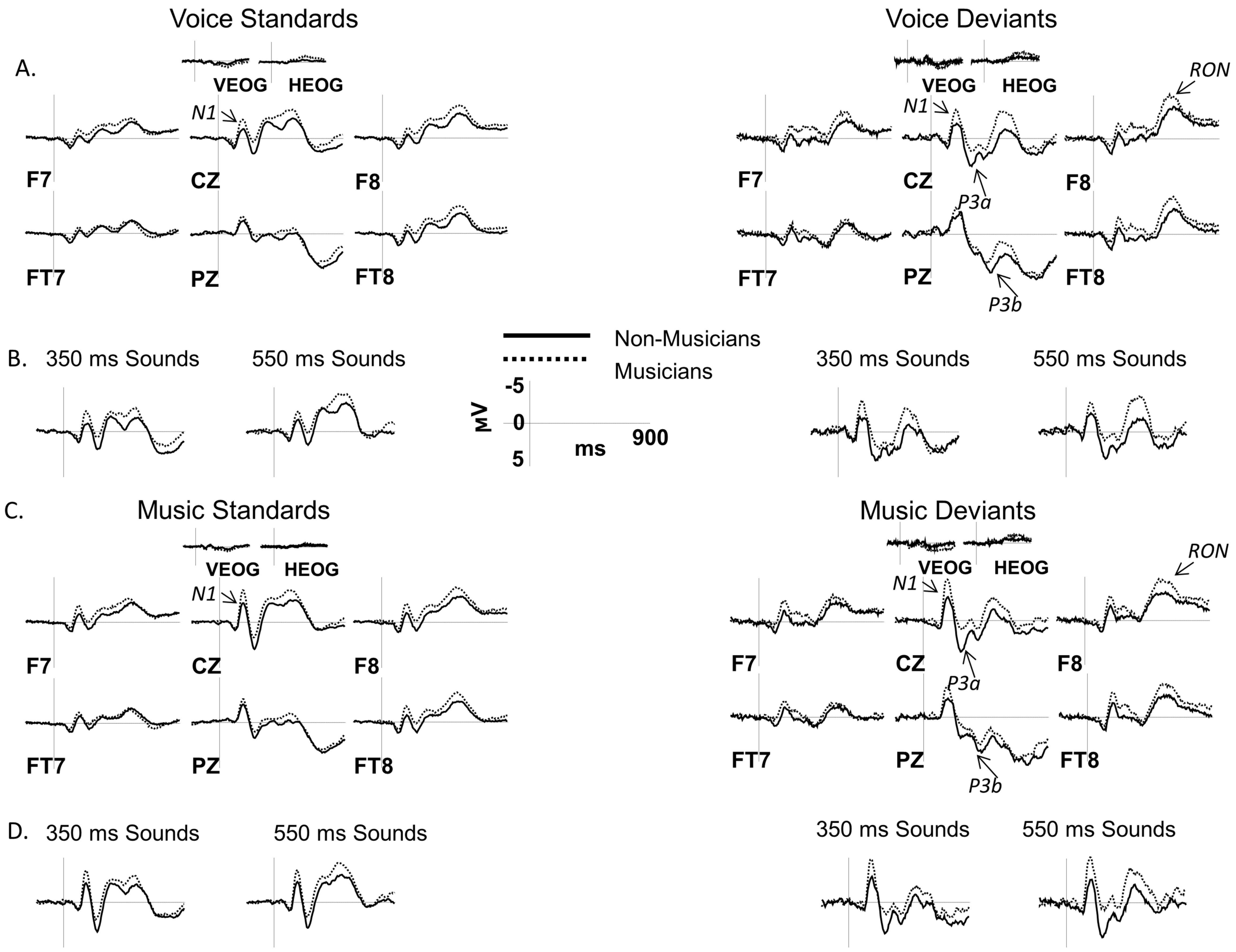

Figure 5.

ERP results in the NAT condition: musicians vs. non-musicians

ERPs elicited by musical and vocal sounds in the NAT condition are shown over a representative set of electrodes. A, C: ERPs from musicians are overlaid with those from non-musicians separately for standards (left side) and deviants (right side) and for vocal (panel A) and musical (panel C) sounds. ERPs were averaged across short and long sound durations for all comparisons. B, D: Separate ERP grand averages for short (350 ms) and long (550 ms) sounds are shown at the CZ site for vocal (panel B) and musical (panel D) sounds. Analyzed ERP components are marked on multiple electrode sites. Negative potentials are plotted upward. Time 0 ms indicates the onset of the stimulus.

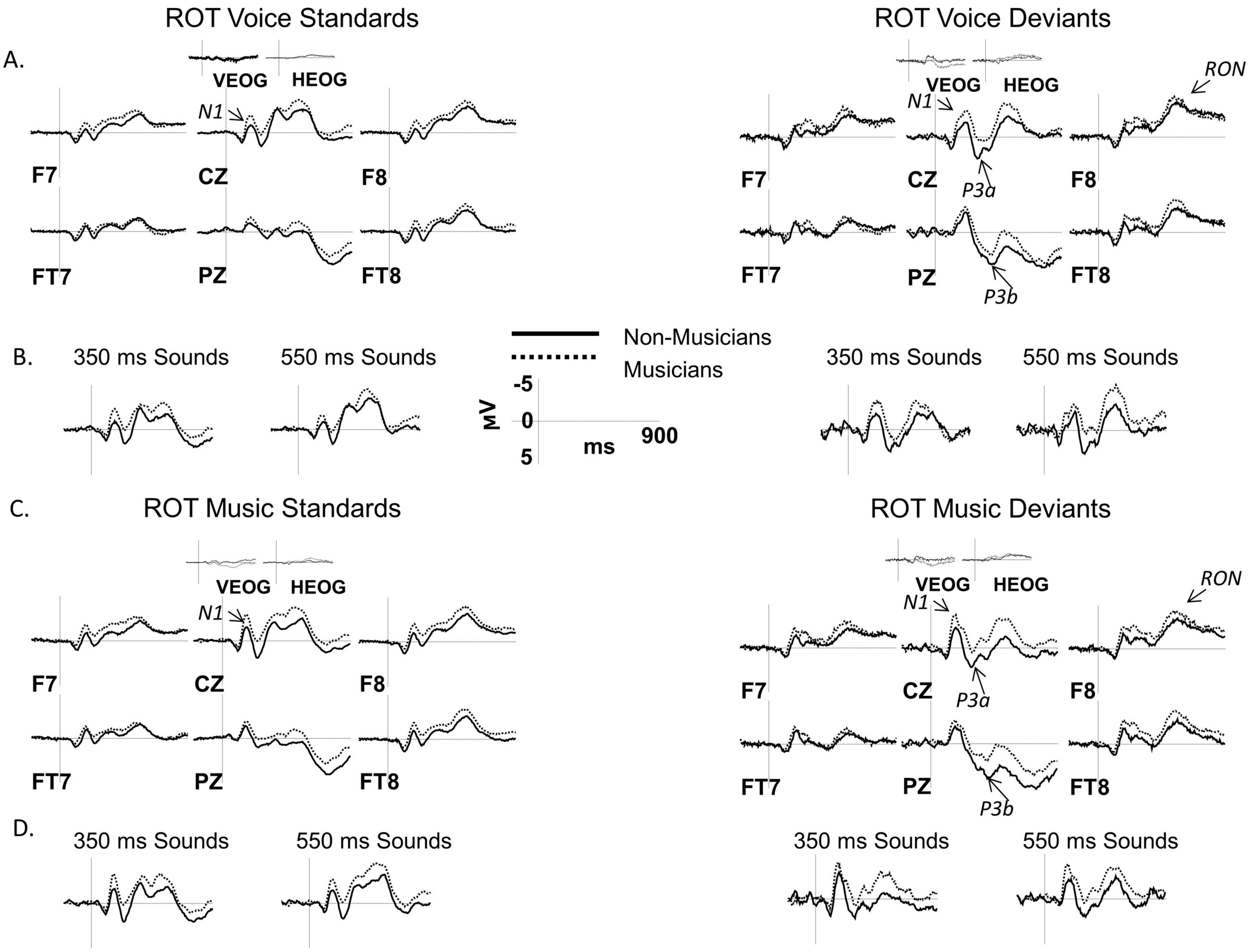

Figure 6.

ERP results in the ROT condition: musicians vs. non-musicians

ERPs elicited by musical and vocal sounds in the ROT condition are shown over a representative set of electrodes. A, C: ERPs from musicians are overlaid with those from non-musicians, separately for standards (left side) and deviants (right side) and for vocal (panel A) and musical (panel C) sounds. ERPs were averaged across short and long sound durations for all comparisons. B, D: Separate ERP grand averages for short (350 ms) and long (550 ms) sounds are shown at the CZ site for vocal (panel B) and musical (panel D) sounds. Analyzed ERP components are marked on multiple electrode sites. Negative potentials are plotted upward. Time 0 ms indicates the onset of the stimulus.

Table 1

Experimental conditions and stimuli.

Music sounds as deviants	
Block 1	
Standard	male voice
Deviants	cello, French Horn
Block 2	
Standard	female voice
Deviants	cello, French Horn

Voice sounds as deviants	
Block 1	
Standard	cello
Deviants	male voice, female voice
Block 2	
Standard	French Horn
Deviants	male voice, female voice

The same block structure was used for NAT and ROT conditions.

Table 2

Temporal windows and electrode sites

Natural sounds condition (NAT)

		Electrode sites		
	Time window (ms)	**midline**	**mid-lateral**	**lateral**
N1	138-222	fz, fcz, cz, cpz, pz	f3/4, fc3/4, c3/4, cp3/4, p3/4	f7/8, ft7/8, t7/8, tp7/8
P3a	222-330	fcz, cz, cpz	none	none
P3b	364-470	cpz, pz	cp3/4, p3/4	tp7/8, p7/8
RON	512-622	fz, fcz, cz	f3/4, fc3/4, c3/4	f7/8, ft7/8, t7/8

Spectrally-rotated sounds condition (ROT)

		Electrode sites		
	Time window (ms)	**midline**	**mid-lateral**	**lateral**
N1	150-234	fz, fcz, cz, cpz, pz	f3/4, fc3/4, c3/4, cp3/4, p3/4	f7/8, ft7/8, t7/8, tp7/8
P3a	222-330	fcz, cz, cpz	none	none
P3b	356-462	cpz, pz	cp3/4, p3/4	tp7/8, p7/8
RON	512-622	fz, fcz, cz	f3/4, fc3/4, c3/4	f7/8, ft7/8, t7/8

Time windows are defined in milliseconds post-stimulus onset.

Table 3

Behavioral results

Accuracy (percent correct)								
	NAT				ROT			
	VS	VD	MS	MD	VS	VD	MS	MD
Musicians	96.4 (1)	94.1 (1.5)	95.5 (1.1)	92.4 (1.7)	94.8 (1.1)	91.7 (2)	95.9 (0.9)	91.7 (1.7)
Non-Musicians	92.9 (1)	91.7 (1.6)	92.7 (1.1)	84.2 (1.8)	91.2 (1.1)	86.2 (2.1)	92.1 (1)	84.4 (1.7)

Reaction Time (ms)								
	NAT				ROT			
	VS	VD	MS	MD	VS	VD	MS	MD
Musicians	831.1 (14.9)	879.2 (15)	837.1 (12.5)	892.1 (16.7)	842.1 (14)	886.5 (17)	828.6 (15.1)	895.8 (15.6)
Non-Musicians	832.2 (15.8)	884.9 (15.8)	836.9 (13.2)	914.2 (17.7)	831.1 (14.8)	894.5 (18)	821.8 (16)	892.3 (16.4)

NAT – naturally-recorded sounds, ROT – spectrally-rotated sounds, VS – voice standards, VD – voice deviants, MS – music standards, MD – music deviants. Numbers in parentheses are standard errors of the mean.