Effects of musical training on sound pattern processing in high-school students

Wenjung Wang\textsuperscript{a,b,*}, Laura Staffaroni\textsuperscript{a}, Errold Reid Jr.\textsuperscript{a}, Mitchell Steinschneider\textsuperscript{c}, and Elyse Sussman\textsuperscript{a,b}

\textsuperscript{a} Dominick P. Purpura Department of Neuroscience, Albert Einstein College of Medicine, Bronx, NY 10461, USA
\textsuperscript{b} Department of Speech, Language, & Hearing Sciences, The Graduate Center, City University of New York, New York, NY 10016, USA
\textsuperscript{c} The Saul R. Korey Department of Neurology, Albert Einstein College of Medicine, Bronx, NY 10461, USA

Abstract

Objective—Recognizing melody in music involves detection of both the pitch intervals and the silence between sequentially presented sounds. This study tested the hypothesis that active musical training in adolescents facilitates the ability to passively detect sequential sound patterns compared to musically non-trained age-matched peers.

Methods—Twenty adolescents, aged 15–18 years, were divided into groups according to their musical training and current experience. A fixed order tone pattern was presented at various stimulus rates while electroencephalogram was recorded. The influence of musical training on passive auditory processing of the sound patterns was assessed using components of event-related brain potentials (ERPs).

Results—The mismatch negativity (MMN) ERP component was elicited in different stimulus onset asynchrony (SOA) conditions in non-musicians than musicians, indicating that musically active adolescents were able to detect sound patterns across longer time intervals than age-matched peers.

Conclusions—Musical training facilitates detection of auditory patterns, allowing the ability to automatically recognize sequential sound patterns over longer time periods than non-musical counterparts.

Keywords
Auditory; Event-related potentials (ERPs); Adolescents; Children; Musicians; Mismatch negativity (MMN)

1. Introduction

Musical training is known to alter sound perception in both children and adults. Those with musical training can detect slight pitch and duration violations in music and speech more accurately \cite{1–3}, and faster \cite{4–6} than non-musicians, and are better than non-musicians at reproducing the order of a three-tone sequence \cite{7}. Effects of musical training are long-lasting,
modulating cortical structure and functions [8–10]. Musicians have higher gray matter volume in motor, auditory, and visual-spatial brain regions compared to non-musicians [11–13], and have been found to have higher concentrations of N-acetylaspartate (NAA) in a brain area associated with music perception (left planum temporale), which correlated with the duration of musical training [14].

The goal of the current study was to evaluate long-lasting effects of musical training in adolescents, by testing the hypothesis that musical training increases the acuity of auditory pattern processing skills, which can be indexed by neurophysiological measures when listeners have no immediate task with the sounds (i.e., passive detection). Perception of musical melody involves detection of temporal sound patterns. Both the pitch intervals and the silence between sequentially presented sounds are encoded. We previously investigated the effect of time on pattern processing as indexed by neurophysiological measures in adults, using a sequentially repeating tone pattern [15]. These results showed that in adults, passive grouping of the tone sequence, indexed by neurophysiologic measures, occurred only when presented at a rapid rate (200 ms stimulus onset asynchrony, SOA). When the same repeating pattern was presented at longer SOAs (400, 600, and 800 ms), there was no neurophysiological indication of the pattern: each tone was processed as an individual unit. Though this study provided some insight in how temporal spacing affected non-musician adults, the question of how musical training affects pattern processing in children has largely been unexplored. Particularly, there is scant literature focused on adolescent abilities in auditory processes. Given the changes in brain structure and function in adult musicians, we hypothesized that active musical training would facilitate the ability to passively detect sequential sound patterns over longer periods of time in adolescents compared to non-musical age-matched peers.

To examine this question, we used the paradigm of Sussman and Gumenyuk [15]. Two tones were presented in a fixed order (XXXXXXXXXXO..., where ‘X’ has a different frequency value than ‘O’) to determine when the five-tone pattern was detected (Fig. 1). The mismatch negativity (MMN) component of event-related potentials (ERPs) was used to physiologically index deviance detection. When the repeating pattern is detected, there is no ‘deviant’ in the sequence and thus, MMN is not elicited by the infrequent ‘O’ tone in the pattern [15–17]. However, when the repeating pattern is not detected, the infrequent ‘O’ tone is detected as an ‘oddball’, and elicits MMN. Thus, in the SOA conditions that MMN is elicited, we conclude there was no passive grouping of the sounds into melodic sequences.

2. Methods

2.1. Participants

Twenty healthy adolescents aged 15–18 years old participated in the study. All participants passed a hearing screening (thresholds ≤20 dB HL for pure tones at 500, 1000, 2000, and 4000 Hz bilaterally) and were paid for their participation ($10 per hour). Participants under 18 years old gave assent and their guardians/parents or participants over 18 years old gave consent for participation. A questionnaire regarding musical training was given prior to testing to determine group placement as ‘musician’ or ‘non-musician’ (Table 1). Ten participants were defined as ‘musicians’ (M = 16 years, 5 months), and 10 as ‘non-musicians’ (M = 17 years). If the participant was actively involved in a music group or classes at the time of testing, then he/she was defined as a ‘musician’. Those who did not participate in any music group or classes were defined as a ‘non-musician’. Thus, all participants had some musical exposure.

2.2. Stimuli

Stimuli were created using Neuroscan software (STIM, Compumedics, Corp., Charlotte, NC). Two pure tones (50 ms duration, 7.5 ms rise/fall time) were presented bilaterally through insert
earphones (E-A-R-tone 3A®) in a fixed order. Sound level (75 dB SPL) was calibrated using a sound level meter (Briel & Kjær 2209). The tone order created a five-tone pattern when repeated (XXXXXOXXXXO... Fig. 1 and [15]). The ‘X’ tone had a frequency of 880 Hz and the ‘O’ tone 988 Hz. The five-tone pattern (XXXXO) was presented 240 times successively, with no breaks between patterns, separately in three conditions of SOA (400, 600, 800 ms). There were 1200 tones presented in each condition.

2.3. Procedures

Participants sat in a comfortable chair in front of a TV monitor during the experiment, and watched a captioned, silent video and had no task with the sounds. The three conditions of SOA (400, 600, 800 ms), presented separately, were counterbalanced across subjects using a Latin Squares design. The total session time was approximately 2 h, which included electrode cap placement and breaks.

2.4. Electroencephalogram (EEG) recording and data reduction

EEG was recorded with an electrode cap from the following electrode sites: Fpz, Fp1, Fp2, Fz, F3, F4, F7, F8, FC1, FC2, FC5, FC6, Cz, C3, C4, CP1, CP2, CP5, CP6, T7, T8, Pz, P3, P4, P7, P8, Oz, O1, O2 (10-20 system) and the left and right mastoids (LM and RM). Fp1 and an electrode placed below left eye were used to monitor the vertical electrooculogram (EOG). An electrode placed on the tip of the nose was the reference and PO9 was used as the ground electrode. EEG and EOG was digitized with a 1000 Hz sampling rate (0.05–200 Hz band pass), and then filtered off-line between 1 and 15 Hz. Epochs were 600 ms starting 100 ms before stimulus onset (–100 to 500 ms). Artifact rejection criterion was set at ±100 μV on all electrodes after baseline correcting to the whole epoch. The ‘standard’ ERP was obtained by averaging together the responses elicited by the 2nd–4th ‘X’ tones of the pattern. The ‘deviant’ ERP was obtained by averaging together the response to the 5th ‘O’ tone of the pattern. The response to the 1st tone of the pattern was excluded from analyses, so not to include refractory effects on responses to the standard following the deviant. MMN was delineated in the difference waveforms, obtained by subtracting the ERP elicited by the standard (‘X’) from the ERP elicited by the deviant (‘O’).

2.5. Data analysis

The peak latencies of the obligatory ERP components (e.g. P1, N1, P2 and N2) were identified using Global Field Power (GFP) analysis [18], and then mean amplitudes were measured using the standard waveform from each condition in a 40 ms window centered on the peak, using electrode sites that provide the greatest signal to noise ratio in adolescents (Fz, Cz, FC1, FC2, C3, and C4; [19]). The MMN peak latency was also identified using GFP, on the grand-mean difference waveform. The mean amplitude was then measured using a 40 ms window centered on the peak at Fz. One-sample, one-tailed t-test was used to determine the presence of MMN. Mixed mode repeated-measures ANOVA with factors of group (musicians vs. non-musicians), condition (400, 600, 800 ms SOA), and electrode (Fz, Cz, FC1, FC2, C3, C4) was used to compare latency and amplitude of the obligatory ERP components. Separate repeated-measures ANOVA with factors of condition (400, 600, 800 ms SOA) and electrode (Fz, Cz, FC1, FC2, C3, C4) were used to compare latency and amplitude of the obligatory ERP components within each group. Greenhouse–Geisser corrections for sphericity were applied and the p values reported. Post hoc analyses were calculated using Tukey HSD.
3. Results

3.1. ERP components

Three obligatory ERP components (P1, P2, and N2) were identified by GFP (See Fig. 2). Latencies and average amplitudes of these components for all conditions are presented in Table 2.

3.1.1. Non-musicians vs. musicians—Fig. 2 displays the ERPs for non-musicians and musicians by stimulus rate. Separate analyses revealed that there was no main effect of group on the amplitude or latency of all obligatory ERP components. There were no significant group differences in the amplitudes or latencies of the P1, P2, or N2 components. Thus, subsidiary within-group analyses were performed using one-way repeated-measures ANOVA to discern effects of SOA on latency and amplitude of the P1, P2, and N2 ERP components.

3.1.2. P1 component—There was a main effect of condition on the P1 amplitude for both non-musicians ($F(2, 18) = 7.73, p < 0.01$) and musicians ($F(2, 18) = 12.10, p < 0.01$). Post hoc analysis indicated that P1 was larger in the 400 ms than the 800 ms SOA condition in non-musicians, and larger than both 600 and 800 ms SOA conditions in musicians (Fig. 3). P1 had a fronto-central distribution (larger at Fz, FC1, and FC2) in both non-musicians and musicians (non-musicians: $F(5, 45) = 43.07, p < 0.01$; musicians: $F(5, 45) = 13.94, p < 0.01$).

P1 latency was significantly shorter in the 800 ms condition than in the 400 ms condition for non-musicians (main effect of condition, $F(2, 18) = 6.53, p < 0.01$), but not for musicians ($F(2, 18) = 0.70, p = 0.94$).

3.1.3. P2 component—P2 was larger in the 400 ms conditions than in the 600 ms condition in non-musicians (main effect of condition for non-musicians, $F(2, 18) = 4.30, p < 0.05$), but there were no significant differences in P2 amplitude by rate in musicians ($F(2, 18) = 0.33, p = 0.72$) (Fig. 3). There was a main effect of electrode site in both groups (non-musicians: $F(5, 45) = 4.28, p < 0.01$; musicians: $F(5, 45) = 18.26, p < 0.01$). P2 was more centrally distributed in non-musicians (largest at Cz), but more fronto-centrally distributed in musicians (larger at Fz, Cz, FC1, FC2). Latency did not significantly change as a function of stimulus rate for either non-musicians ($F(2, 18) = 1.00, p = 0.39$) or musicians ($F(2, 18) = 0.81, p = 0.46$).

3.1.4. N2 component—N2 amplitude did not change as a function of stimulus rate for either non-musicians ($F(5, 45) = 1.71, p = 0.21$) or musicians ($F(5, 45) = 1.41, p = 0.27$) (Fig. 3). N2 was largest over central electrode sites (Cz, C3, and C4) in both groups (non-musicians: $F(5, 45) = 19.02, p < 0.01$; musicians: $F(5, 45) = 24.53, p < 0.01$). Latency was not significantly altered by stimulus rate in either non-musicians ($F(2, 18) = 0.19, p = 0.83$) or musicians ($F(2, 18) = 1.53, p = 0.24$).

3.2. MMN component

In non-musicians, MMN was elicited in the 600 ms ($t_{(9)} = 3.02, p < 0.01$) and 800 ms ($t_{(9)} = 2.42, p < 0.05$) conditions, whereas in musicians, a significant MMN was only elicited in the 800 ms condition ($t_{(9)} = 2.62, p < 0.05$) (Fig. 3 and Table 3). In the 800 ms SOA condition, in which MMN was elicited in both groups, the amplitudes were not significantly different ($t_{(9)} = 2.26, p = 0.82$).

4. Discussion

Adolescents with musical training neurophysiologically group sequential tone patterns over longer time spans than adolescents without musical training (600 ms vs. 400 ms, respectively).
MMN was elicited in the 600 ms SOA condition in ‘non-musicians’, but not in ‘musicians’. The presence of MMN indicates that the pattern was not passively detected: the ‘O’ tone was detected as deviant compared to the ‘X’ tone. Thus, the longest SOA over which the five-tone pattern was detected in the ‘non-musicians’ was 400 ms and the longest SOA for grouping in the ‘musicians’ was 600 ms. These results suggest that musical training alters the physiological response to sound patterns in adolescents. Processing differences were shown for ‘musicians’ who neurophysiologically detected sound patterns over slightly longer time periods than ‘non-musicians’.

These results suggest that musical training facilitates the ability to extract sequential sound patterns over longer time windows than non-trained adolescents. This would allow more flexibility in detecting patterns of incoming sound input, a skill needed to recognize melodies of various time spans. Our results are also consistent with a previous study [20] using similar methodology, and found neurophysiological evidence that adult musicians could detect tone omissions at longer SOAs than non-musicians.

In previous studies, adult musicians have been shown to have larger MMNs than non-musicians [5,21–24]. In contrast, in the current study, there was no significant difference in the size of the MMN between groups. However, in previous studies, adult participants were professional musicians, whereas, in the current study, neither group had professional musicians. There was a considerably smaller difference between our groups in their overall musical training. Thus, our results suggest that even a relatively small amount of musical training can initiate changes in neurophysiological indices of sound pattern processing.

Previous studies have also reported differences in the amplitude of the obligatory responses to sound onsets between musicians and non-musicians for both adults and children. Adult musicians had larger amplitudes for P1 [25,26], N1 [26,27], and P2 [26–29] components. Child ‘musicians’ had larger amplitudes for P1 [30], N1 [30,31], and P2 [30,31] components.

In the current study, the amplitudes of the P1 and N2 components were larger, and the latency of N2 was earlier in the adolescent musicians, but these differences did not reach statistical significance. One explanation is that there was simply not that much of a difference, musically, between our two groups. The non-musicians in our study had received 4 years of musical training on average before they stopped playing instruments. This prior training may already have influenced their neurophysiological responses to sound onsets. Another factor to consider is that the musicians in our study started playing instruments at a later age (around 9-years-old on average) compared to previous studies, which tested children who started playing instruments around 4–6-years-old.

5. Conclusions

Ongoing musical training influences automatic auditory skills related to melody processing. The adolescent ‘musicians’ were more flexible in their ability to detect sequential tone patterns. Musical training facilitates the ability to temporally extract sound patterns over shorter or longer time periods, as would be needed for identifying melody in music.

Acknowledgments

We thank Fakhar Kahn for help with data collection. This research was funded by the National Institutes of Health (R01DC006003).
References


Fig. 1.
Schematic diagram of the stimulus paradigm. The three conditions of stimulus onset asynchrony (SOA, onset-to-onset) are shown: 400 ms SOA (top row), 600 ms SOA (middle row), and 800 ms SOA (bottom row). The ‘X’ represents a tone of 880 Hz, and the ‘O’ represents a tone of 988 Hz. Time is shown on the x-axis and frequency value on the y-axis.
Fig. 2.
Event-related potentials (ERPs) elicited by the ‘X’ tones (the frequently occurring tones) at the Fz electrode site for all three SOA conditions are displayed for musicians (thick, solid line) and non-musicians (thin, solid line). Time (in ms) is represented on the x-axis and amplitude (in μV) on the y-axis. The obligatory ERP components (P1, P2, N2) are labeled.
Fig. 3.
ERPs elicited by both the ‘X’ (thin, solid line) and ‘O’ (thick, solid line) tones and their corresponding difference waveforms obtained by subtracting the ERP elicited by the X-tone from the ERP elicited by the O-tone are displayed for non-musicians (left two columns) and musicians (right two columns) for the three SOA conditions. Time (in ms) is represented on the x-axis and amplitude (in μV) on the y-axis. The obligatory ERP components (P1, P2 and N2) elicited by both stimuli are labeled. Significant MMNs, delineated in the difference waveforms, are denoted with an arrow. MMNs were significantly elicited by the ‘O’ tones in the 600 and 800 ms SOA conditions in the non-musicians, but only in the 800 ms SOA condition in the musicians.
Table 1
Mean (S.D.) and sum of responses to the musical training questionnaire.

<table>
<thead>
<tr>
<th></th>
<th>Non-musicians</th>
<th>Musicians</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>10 (5 M)</td>
<td>10 (5 M)</td>
</tr>
<tr>
<td>Age</td>
<td>17y 0m (7m)</td>
<td>16y 5m (7m)</td>
</tr>
<tr>
<td>Handedness</td>
<td>9 R</td>
<td>10 R</td>
</tr>
<tr>
<td>Age learned to read music</td>
<td>7y 9m (1y 10m)</td>
<td>7y 6m (1y 9m)</td>
</tr>
<tr>
<td>Age began playing</td>
<td>9y 2m (2y 6m)</td>
<td>9y 6m (3y 4m)</td>
</tr>
<tr>
<td>Reads music</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>Number of musician relatives</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Number of years playing</td>
<td>4y 0m (2y 6m)</td>
<td>5y 6m (3y 3m)</td>
</tr>
<tr>
<td>Involved in school music groups</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Involved in outside music groups/classes</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Self-categorization as musician</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>Hours/week spent playing</td>
<td>0.5 h (1 h)</td>
<td>8 h (3 h)</td>
</tr>
<tr>
<td>Hours/week listens to music</td>
<td>15 h (17 h)</td>
<td>32 h (48 h)</td>
</tr>
</tbody>
</table>

M = male, y = year, m = month, R = right.
Table 2

Peak latency (in ms) and amplitude (in μV) of ERP components (P1, P2, and N2) elicited by the frequent X tone in all SOA conditions. Peak latency was identified by Global Field Power (GFP) analysis. NM represents non-musicians and M represents musicians. Amplitude was measured from a 40 ms window centered on the peak latency at Fz for P1 and N2 and at Cz for P2.

<table>
<thead>
<tr>
<th>SOA</th>
<th>P1</th>
<th></th>
<th>P2</th>
<th></th>
<th>N2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NM</td>
<td>M</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Latency (in ms)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>400 ms</td>
<td>75</td>
<td>78</td>
<td>137</td>
<td>130</td>
<td>243</td>
<td>226</td>
</tr>
<tr>
<td>600 ms</td>
<td>68</td>
<td>71</td>
<td>143</td>
<td>139</td>
<td>247</td>
<td>213</td>
</tr>
<tr>
<td>800 ms</td>
<td>64</td>
<td>67</td>
<td>145</td>
<td>141</td>
<td>248</td>
<td>210</td>
</tr>
<tr>
<td>Amplitude (in μV)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>400 ms</td>
<td>1.61</td>
<td>2.51</td>
<td>1.77</td>
<td>1.96</td>
<td>–1.07</td>
<td>–1.47</td>
</tr>
<tr>
<td>600 ms</td>
<td>0.73</td>
<td>1.21</td>
<td>0.85</td>
<td>1.23</td>
<td>–1.74</td>
<td>–1.93</td>
</tr>
<tr>
<td>800 ms</td>
<td>0.46</td>
<td>0.73</td>
<td>1.57</td>
<td>1.48</td>
<td>–1.25</td>
<td>–1.27</td>
</tr>
</tbody>
</table>
### Table 3

Peak latency (in ms) and average amplitude (in μV) of MMN in non-musicians and musicians in three SOA conditions at Fz electrode site. MMN peak latency was identified by GFP analysis from the grand-mean difference waveform. MMN amplitude was measured from a 40 ms window centered on the peak latency at Fz.

<table>
<thead>
<tr>
<th>SOA</th>
<th>Latency</th>
<th></th>
<th></th>
<th>Amplitude</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NM</td>
<td>M</td>
<td>NM</td>
<td>M</td>
<td></td>
</tr>
<tr>
<td>400 ms</td>
<td>157</td>
<td>194</td>
<td>−0.44 (0.86)</td>
<td>−0.31 (0.56)</td>
<td></td>
</tr>
<tr>
<td>600 ms</td>
<td>194</td>
<td>175</td>
<td>−0.94** (0.99)</td>
<td>−0.57 (1.03)</td>
<td></td>
</tr>
<tr>
<td>800 ms</td>
<td>180</td>
<td>202</td>
<td>−1.07* (1.40)</td>
<td>−0.92* (1.11)</td>
<td></td>
</tr>
</tbody>
</table>

* Results of the one-sample, one-tailed t-tests: $P < 0.05$.

** Results of the one-sample, one-tailed t-tests: $P < 0.01$. 