

Published in final edited form as:

Brain Cogn. 2013 November ; 83(2): . doi:10.1016/j.bandc.2013.07.002.

Musical experts recruit action-related neural structures in harmonic anomaly detection: Evidence for embodied cognition in expertise

Jason Sherwin^{1,2} and Paul Sajda¹

¹Department of Biomedical Engineering, Columbia University, New York, NY 10027 USA

²Human Research and Engineering Directorate, U.S. Army Research Laboratory Aberdeen, MD 21001 USA

Abstract

Humans are extremely good at detecting anomalies in sensory input. For example, while listening to a piece of Western-style music, an anomalous key change or an out-of-key pitch is readily apparent, even to the non-musician. In this paper we investigate differences between musical experts and non-experts during musical anomaly detection. Specifically, we analyzed the electroencephalograms (EEG) of five expert cello players and five non-musicians while they listened to excerpts of J.S. Bach's *Prelude* from *Cello Suite No.1*. All subjects were familiar with the piece, though experts also had extensive experience playing the piece. Subjects were told that anomalous musical events (AMEs) could occur at random within the excerpts of the piece and were told to report the number of AMEs after each excerpt. Furthermore, subjects were instructed to remain still while listening to the excerpts and their lack of movement was verified via visual and EEG monitoring. Experts had significantly better behavioral performance (i.e. correctly reporting AME counts) than non-experts, though both groups had mean accuracies greater than 80%. These group differences were also reflected in the EEG correlates of key-change detection post-stimulus, with experts showing more significant, greater magnitude, longer periods of and earlier peaks in condition-discriminating EEG activity than novices. Using the timing of the maximum discriminating neural correlates, we performed source reconstruction and compared significant differences between cellists and non-musicians. We found significant differences that included a slightly right lateralized motor and frontal source distribution. The right lateralized motor activation is consistent with the cortical representation of the left hand – i.e. the hand a cellist would use, while playing, to generate the anomalous key-changes. In general, these results suggest that sensory anomalies detected by experts may in fact be partially a result of an embodied cognition, with a model of the action for generating the anomaly playing a role in its detection.

Keywords

electroencephalography (EEG); expertise; single-trial analysis; pattern recognition; perceptual decision-making

© 2013 Elsevier Inc. All rights reserved.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

The study of the neural processes underlying musical expertise has been an active area of research in cognitive neuroscience. Non-invasive neuroimaging has played an important role in identifying the elements of cognition supporting such expertise. However, the precise roles and relationships of action and perceptual systems remain unclear. In this study, we specifically focus on the perceptual acuity of expert musicians, exploring what role, if any, is played by the interaction of perception and action systems in these subjects. Specifically, we focus on a type of musical expertise that requires a trained mastery of a specific temporal sequence of events, either in the instrumental production of music or in the auditory prediction of a melody, harmony, rhythm and/or timbre. While this may seem a narrow criterion for expertise, we need only to consider the breadth of activities that fall into this classification along with music. For instance, dancing and language comprehension, among many others, share common features with music and have been found to manifest themselves in neural data: pre-motor cortex shows activation via functional magnetic resonance imaging (fMRI) for skilled dancers watching videos of other dancers (Calvo-Merino, Glaser, Grezes, Passingham, & Haggard, 2005; Sevdalis & Keller, 2011); fMRI and electroencephalography (EEG) has shown networks of activation in response to semantic content (Cummings et al., 2006; Gonsalves & Paller, 2000; Hasson, Nusbaum, & Small, 2007; Koelsch et al., 2004; Schmithorst, Holland, & Plante, 2006; Virtue, Haberman, Clancy, Parrish, & Jung Beeman, 2006).

The study of musical expertise has been a highly researched topic. Koelsch, Tervaniemi and others have examined musical experts' pitch and melody processing, showing clear event-related potential (ERP) markers for expertise (Koelsch, Schmidt, & Kansok, 2002; Koelsch, Schroger, & Tervaniemi, 1999). Pfordresher and others have examined the action-related processes of music production using behavioral measures (P. Pfordresher, 2006; P. Pfordresher, Kulpa, JD, 2011). In fact, music perception experiments primarily focus on pitch, along with melody and harmony discrimination, looking at the neural markers (usually ERPs) for deviant tones, notes or chords and the abilities of subjects to recognize syntactically inaccurate musical sequences (Bidelman, Krishnan, & Gandour, 2011; Koelsch, 2009; Koelsch et al., 2004; Koelsch & Siebel, 2005; Loui, Grent-t-Jong, Torpey, & Woldorff, 2005; Maidhof, Vavatzanidis, Prinz, Rieger, & Koelsch, 2010).

It is clear from this previous work that the perception of pitch, melody, harmony and rhythm manifest in measurable neural markers, though the focus has been on markers defined by averaging many EEG trials (e.g., as indexed by the early right-anterior negativity, ERAN) or by analyzing the sluggish blood oxygenation level dependent (BOLD) activity or even by comparing structural connectivity via diffusion tensor imaging (DTI). Here, we aim to investigate the EEG markers for expertise from a different analytic framework, namely through a single-trial analysis of the EEG. Based in statistical pattern recognition and machine learning (L. C. Parra, Spence, Gerson, & Sajda, 2005), this approach is less concerned with the cataloging of particular ERP components (e.g., P300, MMN, ERAN, etc.) and more concerned with – in fact driven by – the distributed EEG activity that discriminates one experimental condition from another. Specifically, we consider the entire electrode space to construct multivariate classifiers and utilize rigorous statistical hypothesis testing in conjunction with signal detection theory (SDT) to determine which electrodes and time points are most discriminating between our chosen experimental conditions. This method contrasts with those of the earlier cited ERP studies, in which electrode regions of interest (ROIs) and post-stimulus time windows are chosen *a priori* for doing statistical testing (usually with ANOVA). Without such *a priori* constraints, the results from these methods would suffer from statistical irrelevance due to multiple comparisons correction.

With this methodological basis, we then determine the differences in discriminating neural activity of a group of musical experts and a corresponding population of novices (i.e., non-musicians). Building from earlier work on auditory-motor interaction in music perception (Zatorre, Chen, & Penhune, 2007), we chose to specifically test the role of action-related processes, especially in experts. We chose an expert subject population with a high degree of proficiency in a particular instrument (cello) and employed a musical stimulus with which they were highly familiar, both in terms of listening to and playing the piece: J.S. Bach's *Prelude* from the *Cello Suite No. 1*. As a control population, we chose a population with no formal music training, nor any experience having played the cello.

In music, an audience can often identify when an expected sequence of events does not occur, i.e., when a 'mistake' occurs. This is the role that a forced key modulation serves as the chosen AME. In this regard, it can bear a strong resemblance to an oddball-like paradigm and in such studies the pool of available subjects with the required level of expertise to perform the task is quite high. For example, the ability to perceive the difference between tones of vastly different frequency relies only on normal hearing amid directed attention (Wronka, Kaiser, & Coenen, 2012) (Chennu & Bekinschtein, 2012). But native speakers of a tonal language (experts) can excel at the task in comparison to non-speakers (novices) when the tones are closer in frequency (Giuliano, 2011; P. Pfordresher, Brown S, 2009). Similarly in music, there is a class of subjects (musical experts) for whom oddball-like stimuli embedded within a particular musical stimulus will evoke a different, and perhaps stronger, neural response than it will in another class of subjects (musical novices). Earlier studies cited above have begun to elucidate this using ERP analysis, but questions still remain regarding the specific roles of action-related processes in such expertise. Even though fMRI highlights the involvement of action-related cortices in experts' perception of music, questions remain as to what role these processes play in anomaly detection (Baumann et al., 2007; Haslinger et al., 2004).

In summary, we designed an oddball-like experiment where both experts and novices (cellists and non-musicians) were instructed to count AMEs, key changes by a semitone (either up or down), that occurred at random in an excerpt. Using this paradigm and our approach for single-trial analysis of EEG, our specific hypotheses were,

1. Experts have greater behavioral accuracy than novices
2. Experts have a more pronounced and rapid neural response to AMEs relative to novices
3. Experts utilize neural machinery for detecting AMEs that is reflective of their extensive instrumental and musical training.

2. Materials and Methods

2.1 Subjects

Ten subjects participated in the study, five were classified as experts (3 females, 2 males) and five were novices (2 females, 3 males). The size of our expert population was limited by the number of concert-level cellists we were able to recruit for the study. Despite this limited number of subjects, we found statistically significant results to test our aforementioned hypotheses relative to group effects.

The ages and years of formal music and cello experience are given in Table 1. Experts had 32.6 ± 10.0 years of cello performance training and had a mean age of 41.2 ± 10.6 years. All experts were professionals, played the cello with their right-hand and had played the J.S. Bach piece extensively as part of training and performing. The novices had a mean age of

30.8±5.8 years. Though novices were vaguely familiar with the piece used in the study, they had neither extensive knowledge of the piece, J.S. Bach's music nor the specific recording. They also had no prior experience playing the cello nor any formal musical training (beyond taking a music class in high-school). Furthermore, the novices were not significantly different in age ($p>0.05$, independent groups t-test) or gender from the *experts*. All subjects reported normal hearing and no history of neurological problems. Informed consent was obtained from all participants in accordance with the guidelines and approval of the Columbia University Institutional Review Board.

2.2 Stimuli and Behavioral Paradigm

We used the first 65 seconds of Yo Yo Ma's recording of the *Prelude* of J.S. Bach's *Cello Suite No. 1* as our stimulus. Subjects listened to this complete musical section, or excerpt, 40 times, with 32 of these excerpts altered to contain 3 to 6 anomalous key-change events. There were 140 total key-change events in the experiment.

Anomalous key-shifts were added to the original 44.1kHz .mp3 file using Apple Inc.'s Logic Express 9.0 (Cupertino, CA). These key-changes were inserted at random times with Pitch-Shifter, a built-in plug-in to Logic Express, and each trial was then saved as a 44.1kHz .wav file. Although the pitch of the .wav file is raised or lowered with this algorithm, the effect is a complete change of musical key on the sound file.

Neither the frequency, timing, nor direction, of key-changes could be predicted from one another. The rule for the key-changes was that the net key-change should be no more than a half step from the original key of the recording. In the case of the Bach prelude (originally in G), the recorded key was altered no higher than G-sharp (G#) and no lower than G-flat (Gb). This was done to avoid overt distortion of the original recording, thereby making a key-change obvious from non-harmonic considerations (e.g., timbre). A schematic representation of a key-change trial beginning and ending in the original recording key is shown in Figure 1. In the figure, a blue step-function above the musical notation provides a schematic of the key for those not fluent in musical notation.

We balanced the direction of key-changes using a total of 69 key-changes up and 71 key-changes down. Considering each key-change as a stimulus, the inter-stimulus-interval (ISI) was at least 6 seconds. To remove the potential for bias, especially among our experts, some trials began one semitone down or up from the original key to ensure that the task was specific to relative key-change, rather than deviation from absolute key.

The 8 control and 32 key-changed excerpts were presented to the subjects in a pseudorandom order across eight blocks, each containing five excerpts. Each block contained either none or at most two control excerpts in which no key-changes occurred. An example presentation order is shown in the Appendix.

A Dell Precision 530 Workstation was used to present the audio stimuli with E-Prime 2.0 (Sharpsburg, PA) and a stereo audio card. The subjects sat in an RF-shielded room between two Harmon & Kardon computer speakers (HK695-01, Northridge, CA) connected to the Dell. Each subject was allowed to adjust the volume so that they could comfortably perform the task (playback volume did not exceed 80dB).

Subjects performed a simple detection task in which they were asked to respond covertly by counting the number of AMEs they heard. Counting was used rather than an overt behavioral response, such as a button-press, to minimize motor confounds in the EEG. Counts provided an estimate of task performance, thereby ensuring that subjects attended to the task. Subjects were not explicitly instructed to detect a key-change, but simply to pay

attention for anything out of the ordinary—i.e. an anomaly. They were instructed not to move during the task and they were monitored for any movement both visually and via analysis of motor-related artifacts in the EEG. Stimulus events were passed to the EEG recording through a TTL pulse in the event channel. In post-hoc analysis, stimulus events were added to the EEG of control tracks at the times when they had occurred in the key-change tracks.

2.3 Data Acquisition

EEG data was acquired in an electrostatically shielded room (ETS-Lindgren, Glendale Heights, IL, USA) using a BioSemi Active Two AD Box ADC-12 (BioSemi, The Netherlands) amplifier from 64 active scalp electrodes. Data were sampled at 2048 Hz. A software-based 0.5 Hz high pass filter was used to remove DC drifts and 60 and 120 Hz (harmonic) notch filters were applied to minimize line noise artifacts. These filters were designed to be linear-phase to minimize distortions. Stimulus events – specifically, key-changes – were recorded on separate channels.

Throughout the experiment, subjects listened to the music excerpts with eyes closed. This minimized blinks and eye-movement artifacts. This technique has been used in other music perception studies (Maidhof et al., 2010), as well as auditory oddball studies (Goldman et al., 2009). Consequently, no eye calibration experiments were needed before implementing the filtering described above.

In epoching the data, the average baseline was removed from 1000ms pre-stimulus, i.e., 1000ms before the AME or its corresponding control time in the epoch. After epoching into stimulus and time-matched control events, an automatic artifact epoch rejection algorithm in EEGLAB (Delorme & Makeig, 2004) was run to remove all epochs that exceeded a probability threshold of 5 standard deviations from the average.

2.4 Data Analysis

2.4.1 Behavioral Accuracy from Post-Excerpt Reporting—We calculated behavioral accuracy based on a post-excerpt reporting of how many AMEs the subjects counted. Subjects reported this number after each of the 40 listening excerpts. We calculated accuracy by noting the deviation from the actual number of key-changes in each excerpt. For instance, if excerpt i contained n_i key-changes and a subject reported $n_i - k_i$ or $n_i + k_i$ key-changes then this constituted an error of k_i . However, if the subject reported n_i then this constituted an error of zero, $k_i = 0$. To summarize the performance of each subject, we subtracted from 1.00 (perfect accuracy) the total number of errors, normalizing by the total number of actual key-changes that occurred (Equation 1). Despite the possibility of accuracy being less than zero, this case would only be in the event of extremely poor behavioral performance and did not occur in our experiment.

$$Accuracy = 1 - \frac{\sum_{i=1}^{40} k_i}{\sum_{i=1}^{40} n_i} \quad \text{Equation 1: Behavioral Accuracy Calculation}$$

2.4.2 Single-trial Analysis of the EEG—We performed a single-trial analysis of the filtered, epoched and artifact-removed EEG to discriminate between anomalous key-changes in either direction, regardless of starting and ending key, and their corresponding control epochs. With the lack of an overt response to the key-change, we necessarily included both hits and misses in these epochs, thereby making the discrimination challenging if the subject

does not demonstrate sufficient behavioral performance on the task. Logistic regression was used to find an optimal projection in the EEG sensor space for discriminating between these two conditions over each sub-window of the entire epoch (L. Parra et al., 2002; L. C. Parra et al., 2005). Specifically, we defined a training window starting at either a pre-stimulus or post-stimulus onset time τ , with a duration of δ , and used logistic regression to estimate a spatial weighting vector $\mathbf{w}_{\tau,\delta}$ which maximally discriminates between sensor array signals X for each condition (e.g., key-changes versus controls). For our experiments, the duration of the training window (δ) was 50ms and the window onset time (τ) was varied across time $\tau \in [-200,950]$ ms in 25ms steps (50% overlap), thereby covering $[-200,1000]$ ms. This training window size and overlap has been successfully used in other implementations of this technique (L. Parra et al., 2002; L. C. Parra et al., 2005), as it allows a suitable balance between local and global temporal EEG dynamics. We used the re-weighted least squares algorithm to learn the optimal discriminating spatial weighting vector $\mathbf{w}_{\tau,\delta}$ (Jordan, 1994).

$$\mathbf{y} = \mathbf{w}_{\tau,\delta} X \quad \text{Equation 2: Projection Equation for Component}$$

The result is a ‘discriminating component’ \mathbf{y} that is specific to activity correlated with each condition while minimizing activity correlated with both task conditions such as early audio processing. The term ‘component’ is used instead of ‘source’ to make it clear that this is a projection of all activity correlated with the underlying source. In Equation 2, X is an $N \times T$ matrix (N sensors and T time samples).

Once solving for optimal discriminating spatial vectors in each window we can compute the electrical coupling coefficients (Equation 3).

$$\mathbf{a} = \frac{X\mathbf{y}}{\mathbf{y} \bullet \mathbf{y}} \quad \text{Equation 3: Sensor Projection Onto Discriminating Component}$$

This equation describes the electrical coupling \mathbf{a} of the discriminating component \mathbf{y} that explains most of the activity X . Since \mathbf{a} is in the sensor space, we can use it to obtain a topological map of which electrodes discriminate the most for each condition.

We calculated the ‘EEG image’ by applying $\mathbf{w}_{\tau,\delta}$ to the EEG data of each window ($X(\tau')$), $\tau' \in [-200,950]$ ms. Given a fixed value of τ , the result of this calculation provides a trial-by-trial visual representation of the window during which the discriminating component is at its highest value (see Results).

We quantified the performance of the linear discriminator by the area under the receiver operator characteristic (ROC) curve, referred to as A_z , with a leave-one-out approach (Duda, 2001). The ROC is a curve of false positive rate vs. true positive rate, therefore greater area values under this curve indicate more accurate classification. We used the ROC A_z metric to characterize the discrimination performance between key-change and corresponding control epochs while sliding our 50ms training window from start times of -200 ms to 950 ms post-stimulus (i.e., varying τ). This epoch size provided substantial time both before and after the stimulus to observe any possible neural correlate of an anticipation of the AME.

We quantified the statistical significance of A_z in each window (τ) via a permutation-based relabeling procedure. In particular, we randomized the truth labels between control and key-change epochs and retrained the classifier. This was done 250 times for each subject in each of the forty-seven 50ms windows, yielding 11500 permutations for each subject. On a group level (10 subjects), this yields 115000 permutations. On a subject-level significance analysis, we utilized the false discovery rate (Benjamini, 1995) at $p = 0.05$, unless otherwise

specified. On a group-level significance analysis, we utilized the Bonferroni correction at $p = 0.05$ (i.e., $p = 0.05/47 = 0.001$). For both levels, the number of permutations provided a suitable distribution to gauge statistical significance, regardless of the number of multiple comparisons in epoch-time.

2.4.3 Traditional ERP Analysis—We also performed a traditional evoked-response potential (ERP) analysis of the filtered, epoched and artifact-removed EEG. We did not consider *a priori* scalp regions of interest (ROIs), peaks and/or times as others have done (e.g., (Koelsch et al., 2002; Koelsch et al., 1999; Loui et al., 2005; Maidhof et al., 2010)). Rather, we utilized the statistical significance of our single-trial analysis after correcting for multiple comparisons to determine which windows were most significant (i.e., max significant A_z). This approach is similar to following the peak activity of a component (e.g., P3, N2, etc.), but has the added benefit of not needing to specify ROIs *a priori* since the peak discriminating activity is across the whole scalp. The ERPs from these subject-specific times were then used to consider grand averages within and between subject groups, as well as between different key-change events (up or down).

3. Results

3.1 Behavioral Performance Shows Experts Out-perform Novices

Without explicit instruction to detect key-changes all subjects, regardless of their expertise, were able to perform the task with at least 80% accuracy, thereby demonstrating the saliency of a forced key-change even to the novice listener. From the behavioral data summarized in Figure 2 (right bars), we see, however, that novices have a significantly lower accuracy rate than the experts (accuracy of 0.81 ± 0.04 vs. 0.94 ± 0.01 , $p < 0.02$, independent groups t-test).

We also examined the dependence of accuracy on experiment time by considering the Pearson correlation between errors and block number. Once applying a Bonferroni correction for independent multiple subject comparisons, there were no subjects that showed significant correlation between errors and block number ($p > 0.19$), indicating that behavioral performance did not change significantly as a function of experiment time. This result lends further evidence to the saliency of the key-change event to both groups from the very beginning of the experiment.

3.2 Single-trial Analysis Reveals Differences in Neural Activity Between Novices and Experts

Using the sliding window logistic regression classifier (see Methods), we found only post-stimulus windows of significant discrimination for each group of subjects. Figure 3 shows each group's mean discrimination vs. epoch time. On average the leave-one-out (LOO) discrimination at each window is substantially greater for experts than it is for novices ($p < 0.01$, paired t-test). From visual inspection we see that there are more discriminating windows for experts, as well as higher peak discrimination.

To further quantify the differences between groups, we can examine subject-level LOO results. Setting the false discovery rate (FDR) for each subject to $p = 0.01$, we find that experts have more significant discriminating windows than do novices ($p < 0.02$, independent groups t-test). Table 2 shows the values and times of maximum discrimination (A_z) for each subject. All discrimination values and corresponding times in this table are FDR corrected at a significance of $p < 0.05$. Experts (325 ± 30 ms) are faster to their maximum A_z than novices (550 ± 75 ms) ($p < 0.02$, independent groups t-test). Experts also exhibit higher values of maximum A_z than do novices ($p < 0.01$, independent groups t-test), the latter of which can be seen in the behavioral accuracy differences in Figure 2.

Finally, we utilized the discriminator's output to examine the single-trial variability of each subject to the key-change. We filtered each epoch's sensor data (X) with the classifier ($w_{\tau,\delta}$) that yielded the maximum value of discrimination (y) between key-changes and controls (i.e., such that τ is of maximum A_Z). This allowed us to examine the strength of the discriminating component, as well as its trial-to-trial variability. Generally, we found a common window in key-change trials across which each subject showed high values of discriminating activity compared both to neighboring windows and corresponding control trials' windows. We also found differences between experts and novices. For instance, Figure 4 shows the discriminator outputs (mean of y in Equation 2 across the window) of an age- and gender-matched novice and expert. From this figure, we see the demonstrated timing difference of maximum A_Z between novices (Figure 4A) and experts (Figure 4B).

Furthermore, we see that the discriminating component for the expert during key-change trials is generally greater than that of the novice at the window of maximum discrimination. This observation also extends to the group level, where we find that the mean of the window-measured $y_{\max A_Z}$ during key-change trials is greater ($p < 0.05$, independent groups t-test) for experts ($2.72 \pm 0.14 \mu V$) than it is for novices ($2.22 \pm 0.23 \mu V$), indicating stronger discriminating activity among the experts at peak neural response to the key-change.

3.3 Traditional ERP Analysis Shows Group Differences at Peak Discrimination

We utilized our discriminator ($w_{\tau,\delta}$) to examine the grand average ERP of each group at subject-specific times of peak discrimination. Since each time is actually the start time of a 50ms window, we averaged across time within this window of peak A_Z . We also considered the differences, if any, between up and down key-changes for either group. After epoch rejection, we found no significant differences between up (64 ± 1) and down (65 ± 1) trials for any subject ($p > 0.10$, paired t-test). With this balance in the two types of key-change events, we calculated grand average ERPs for both experts and novices during up and down events. Figure 5 shows no obvious difference, within group, between key-change types (up or down) and we verified this with a two-sample Kolmogorov-Smirnov test in sensor space ($p_{\text{experts}} > 0.90$ and $p_{\text{novices}} > 0.90$). Furthermore, the novices exhibit a posterior P300 (Bledowski, Prvulovic, Goebel, Zanella, & Linden, 2004; Bledowski, Prvulovic, Hoechstetter, et al., 2004), whereas the experts' activity is frontal. Importantly, in comparison to previous work (e.g., (Koelsch, 2009; Koelsch et al., 2002; Koelsch et al., 1999)), we do not find the ERAN to be the most discriminating ERP component for musical experts, as their peak discrimination occurs at 325 ± 30 ms, which is after the traditional time of the ERAN. In addition, the scalp distribution is substantially different from the classic ERAN. Among experts, we do find a strong posterior negativity at peak discrimination, but it is later than the traditional timing of the ERAN. From its later timing, it is likely that this negative component is related to semantic processing and the N400 (Koelsch et al., 2004).

3.4 Forward Models of Discriminating Components Show Classic P300 Topologies in Novices But Not Experts

To more closely examine the differences between experts and novices in terms of the spatial distribution of their maximally discriminating components, we estimated the electrical coupling, \mathbf{a} , (i.e., the forward model) for each group. Figure 6 shows these forward models for experts and novices computed using the components at each subject's maximum A_Z . Specifically the forward model for each subject-class represents an average of the subject-specific forward models estimated using the window of maximum A_Z (i.e. the times in Table 2 represent the τ 's for estimating the components in Equation 2 and the resulting forward models \mathbf{a} using Equation 3).

Clear from Figure 6 is a difference in the forward models of the discriminating components for the expert and novice groups. This difference is consistent with the ERP results shown in Figure 5A/B and Figure 5C/D. The novice group (Figure 6A) has a topology consistent with the posterior P300 (Bledowski, Prvulovic, Goebel, et al., 2004; Bledowski, Prvulovic, Hoechstetter, et al., 2004). As with the ERP results, this is consistent with what one might expect for a target versus distractor task if we consider the key-changes as targets embedded in a stream of distractors (the ongoing musical piece). The topology for the experts looks quite different, with Figure 6B showing a strong activation of frontal sites and a corresponding deactivation of occipital sites for the target condition. Such a topology is more consistent with neural activity seen in trained instrumentalists following a performance error with audio and motor feedback (Ruiz, Jabusch, & Altenmuller, 2009).

3.5 Source Modeling Indicates That Experts Recruit Motor Related Areas for Anomaly Detection

We used source estimation to examine the cortical generators of the differences seen between experts and novices at the scalp level. Low-resolution tomography (sLORETA) of scalp potentials has been extensively employed to find possible cortical origins of such activity (Pascual-Marqui, 2002; Pascual-Marqui, Esslen, Kochi, & Lehmann, 2002) and so we computed the sLORETA estimates of the neuronal current source distributions.

To compare experts and novices, we grouped up and down key-change trials, given our earlier results showed that there were no significant differences between those anomaly types. With this increased statistical power we used statistical non-parametric mapping (SnPM) for comparing experts and novices in the sLORETA voxel space. We calculated sLORETA fits for each subject's grand average ERP at the subject specific maximum A_z for up and down epochs, respectively. As for the ERP calculations, these are averaged across time in the 50ms window of peak discrimination between key-changes and controls. We then compared the ten novice sLORETA fits with the ten corresponding expert fits. The sLORETA parameters used in the independent groups t-test can be found in the Appendix. We established significance using 1024 permutations and the SnPM procedure for voxel-space comparisons (Holmes, Blair, Watson, & Ford, 1996; Nichols & Holmes, 2002; Pascual-Marqui, 2002; Pascual-Marqui et al., 1999). Figure 7 shows the t-distribution values of the log of the ratio of averages (similar to a one-way ANOVA, $F(1,19)$) for experts > novices (purple/blue) and novices > experts (orange/yellow) mapped to a six-view projection of the cortex. We find that experts have 15 voxels across BA 6 and BA 9 with significantly greater activity than novices ($p < 0.01$, independent groups t-test), peaking at MNI (20, -10, 60). Interestingly, this voxel of peak activity has also been implicated in music imagery tasks in pianists (Baumann et al., 2007), though in that work bilateral activation was found in this part of the premotor dorsal cortex. For our results, not only is peak activity for experts in right motor cortex, but we also find more right than left lateralized activation when we consider SnPM-corrected voxels out to $p < 0.05$ (31 left voxels vs. 50 right voxels, 17 of which are in right frontal cortices). All 31 left voxels are in the motor cortices, while right voxels are distributed between the right motor (33 voxels) and frontal cortices (17 voxels). Table 3 gives MNI coordinates for a subset of these voxels with their t-distribution values (thresholded by $p < 0.01$). There are no voxels showing significance for novice activity greater than experts ($p = 0.73$).

We also tested the time-uniqueness of this response amongst experts by considering windows of non-maximum A_z . Such a test addresses possible concerns that the motor response seen in the experts (cellists) is only from them listening to their instrument of expertise being played (Zatorre et al., 2007), rather than being an additional neural correlate of the key-change detection event. To this end, we randomly selected a pre-stimulus window

(i.e., $\tau \in [-1000, -50]ms$) for each subject, rather than the window of maximum A_z , and performed the same sLORETA statistical analysis as before. We did the same calculation for a randomly chosen post-stimulus (i.e., $\tau \in [0, 950]ms$) FDR-corrected insignificant and FDR-corrected significant window (non-maximum A_z), respectively. For these three randomly selected window cases, we found no voxels having significantly greater activation for experts or novices ($p > 0.05$, independent groups t-test, all three cases).

We also tested the robustness of our result among experts by averaging in the voxel space, rather than the sensor space, before statistical testing. Due to sLORETA's guarantee of zero-error in fitting the sensor distribution to the voxel space, and linearity of electromagnetic sources, we expected to duplicate our earlier findings. We transformed each epoch's time-averaged window of subject-specific maximum A_z into voxel space using sLORETA (ups and downs separately). We then averaged within subject all epochs for key-changes up and down, respectively. Performing the same SnPM and f-test, we found peak activity once again at MNI (20, -10, 60) and the same 15 voxels showing greater activity among experts than novices ($p < 0.01$, independent groups t-test).

Finally, the slight right lateralization of the neuronal current response among experts and the frontal source found from this analysis is consistent with both the ERP and forward model results, Figure 5B/D and Figure 6B, respectively.

4. Discussion

In this paper, we have shown that both the timing and sources of discriminative neural markers are different between our expert and novice subjects in an oddball-like musical anomaly detection task. Without particular instruction to detect key-changes, all subjects detected the AMEs, with group differences manifesting themselves in behavioral performance, discriminating neural activity, traditional ERP analysis, scalp topology of discriminating component forward models and the distribution of neuronal sources. We now discuss these results in the context of relevant and related studies.

4.1 Studying Expertise in Musicality

Other studies have investigated the neural correlates of experts and novices with respect to musical stimuli. For instance, expert pianists were shown to have less fMRI activation than control subjects in pre-motor cortex during complex movement tasks at a piano keyboard, indicating a learning effect (Meister et al., 2005); pianists have been shown to have higher fMRI activation in motor areas when listening to musical stimuli than non-pianists (Baumann et al., 2007); EEG has been used to show neural signatures that precede when a trained pianist is about to hit an incorrect note (Ruiz et al., 2009). EEG also has been used to examine the role of auditory feedback in trained vs. untrained pianists, where it was found that an N210 ERP was seen for experienced vs. less-experienced pianists following an alteration of the auditory feedback (Katahira, Abla, Masuda, & Okanoya, 2008). Many EEG studies also have investigated augmented pitch processing capabilities of expert musicians (Koelsch et al., 2002; Koelsch et al., 1999).

While these previous studies provide hints of neural markers of musical expertise, their conclusions stem from the results of neural activity measured over periods of time that are long when compared to the underlying neural dynamics. For instance, the temporal resolution of fMRI is typically constrained by the repetition time TR and the sluggish BOLD response. Traditional EEG studies on the other hand, while they reveal phenomena with high temporal resolution, are mostly the result of averaging over many trials that unfold in time across many minutes, ignoring the variability of the activity across trials.

Nevertheless, we can directly compare our findings to the existing EEG literature on music perception and expertise, at least where the experimental task is similar. Maidhof et al. (Maidhof et al., 2010) employ a task, similar to ours, in an attempt to separate the action-and-perception-related processes in musical deviance, i.e., AME, detection. Since there is no action required by our subjects, the closest analog to our experimental design is from the 'perception condition' of this paper. Maidhof et al. however define an AME differently than we do in our experimental design. Specifically, in their paper, an AME is a single note that is flattened by one semitone from its diatonic scalar context, whereas our study's AME is a semitone key-change that does not return to the previous musical context after the event. Comparing our experts' EEG activity with those of Maidhof et al., we find comparable grand average ERPs in *a priori* ROIs (such as Fz, FCz and Cz) and scalp distributions in *a priori* post-AME time windows (e.g., 180–220ms, 270–320ms and 370–430ms). These figures can be found in Appendix for direct comparison to Figure 3 of Maidhof et al. The primary differences between our ERPs and those of Maidhof et al. are that our N2 and P3 complexes are smaller in magnitude, though their timings are comparable. Our ERPs also exhibit a later negativity, most notable in Fz and FCz. These differences are likely due to the nature of the AME used in our experiment. In Maidhof et al., the AME lasts 104–217ms but the harmonic context is quickly re-established. This is not the case in our experiment. Rather, we change key by forced, instead of diatonic, modulation. Consequently, the long-term harmonic expectations established from the larger musical context preceding the AME are not fulfilled. These contextual or semantic-related responses have been seen in music in the N400 and N500 (Koelsch, 2009; Koelsch et al., 2004; Loui et al., 2005) and in language in the N600 and P820 (Cummings et al., 2006; Gonsalves & Paller, 2000).

Other studies have specifically examined particular ERP components, such as the ERAN or MMN. Koelsch has looked at these components extensively in musical expertise (Koelsch, 2009; Koelsch et al., 2002). Although not a particular aim of our study, we can consider our results in the context of Koelsch's studies of the ERAN and MMN. We particularly focus on the ERAN because it is more dependent on music-syntactic regularities extant in long-term memory than the MMN, which depends on online establishment of regularities in auditory stimuli and is therefore not specific to music. We find concordant results with Koelsch (Koelsch, 2009) and Loui et al. (Loui et al., 2005) in that the ERAN of the experts is larger than the novices, especially in Fz and Cz (not shown).

Although we corroborate earlier results on musical expertise, we see some differences that are likely due to the differences in task. The most obvious difference between groups is in the P3 amplitude, which is much larger for experts than novices in both electrode ROIs. Furthermore, the latency of the experts' P3 is smaller than that of the novices at Cz and is likely due to their higher sensitivity to the AME. We additionally find, through single-trial analysis, that cellists peak in neural discrimination earlier than do the novices. Whereas previous ERP studies rely on electrode ROIs and time windows, we determine which time windows and electrodes are maximally discriminative between key-change and control events. Finally, we found a larger late negativity (~600–700ms) in the cellists than non-cellists (not shown). This late negativity is not seen in Koelsch (Koelsch, 2009), presumably because such late negativities are also associated with semantic content developed over longer periods of time (Cummings et al., 2006; Gonsalves & Paller, 2000; Koelsch, 2009; Koelsch et al., 2004; Loui et al., 2005) than the stimuli they used. Compared to only 2.5s of preceding musical context in Koelsch, we utilized at least 6s of pre- and post-AME musical context. Therefore, it is possible that we see a late negativity in both groups (accentuated in cellists) because of the semantic content conveyed by the harmonic context.

Our work differs from previous work using ERP analysis because we were not constrained by electrode ROIs nor pre-determined windows of EEG activity that we can see are heavily

dependent on the stimuli and tasks (e.g., compare Maidhof's results (Maidhof et al., 2010) to ours in the Appendix). Our experimental design focused on creating a stimulus salient to both groups. We found that the expertise of the professional cellists allowed them to more accurately and more quickly identify the anomalous events in a sequential stream of musical stimuli. Therefore, a primary contribution of our paradigm and ensuing analysis is that it has focused on the neural markers of AME detection that varied trial-to-trial (see Figure 4). We utilize this variability to identify discriminating EEG components and the times of the maximum discriminating components (Figure 3 and Table 2), which ultimately inform our ERP (Figure 5), forward model (Figure 6) and source localization (Figure 7) analyses. Consequently, while our results have been demonstrated in the context of musicality, they can theoretically be extended to other domains of human interaction with the external environment that depend on a tight coupling of sensorimotor interaction to cognition (e.g., language comprehension and production).

4.2 Auditory-Motor Coupling in Experts

Our source localization results reveal strong activity amongst experts in right lateralized frontal and bilateral motor, supplementary motor and movement-related cortices (Figure 7), though the activity is more right than left lateralized. Significant values from the comparison statistic ($p < 0.01$) are found at locations implicated in movement(s) and timing experiments involving the dorsal pre-motor cortex (dPMC) and supplementary motor area (SMA).

In literature specific to music, there is a strong connection emerging between audio and motor processing. Zatorre et al. (Zatorre, 2007; Zatorre et al., 2007) present a literature survey in which the case is made for such a connection. They specifically focus on dorsal PMC (dPMC) as the location of mirror (or echo) neurons that link auditory and motor systems. Our source localization is consistent with this argument as we have also found an audio-motor connection in the dPMC. They also claim that dPMC is involved in the higher-order feature extraction of an auditory stimulus to implement temporally organized actions. In our experiment, the higher-order feature would be the key of the recording, so there appears to be even more concord with this proposed role of the dPMC.

Our study showed a strong activation in motor areas for the expert cellists, though it is possible that this difference is due to over-familiarity with the piece. Leaver et al. (Leaver, Van Lare, Zielinski, Halpern, & Rauschecker, 2009) found that non-musically-trained subjects showed stronger activations in fMRI of pre-SMA and ventral PMC when listening to highly familiar vs. unfamiliar melodies. In our findings, the basis of the stimulus set was chosen in particular for its familiarity both to novices and experts. Its familiarity to novices was based more on a subjective notion of the prelude's permeation in popular culture, rather than a quantitative measure of exposure. The experts on the other hand are highly familiar with this piece, not only from hearing it, but also from playing it many times since the early days of their training. Thus the connection we found to SMA is not surprising in the context of experts' familiarity to the piece being a prerequisite. Furthermore, the greater right-lateralized response in experts centered on motor cortices associated with movement of the left hand, wrist and arm provide a strong case for the familiarity extending into the motor domain.

The use of anomaly detection, as opposed to listening straight through, sheds light on how a deviation from the expected sequence manifests itself in an expert performer's brain. Haslinger et al. (Haslinger et al., 2004) compared pianists and non-pianists using fMRI in a bimanual finger tapping task and found greater activation in pre-SMA and rostral dPMC for non-pianists during task execution. The authors conclude from this study that expert pianists required fewer neural resources to execute the task due to their years of training. In the case of our target detection paradigm though, we are introducing an error into the musical

sequence. If echo neurons were responsible for the auditory-motor interaction described by Zatorre and others, then our findings of neuronal source activity in SMA would mean that experts respond to the key-change as an unexpected (i.e., untrained) motor movement. The lack of such a response when we choose pre-stimulus or otherwise non-maximally discriminating windows post-stimulus reinforces this role for SMA in key-change detection. The greater right lateralization of the activity in cellists further strengthens this claim, since the pitch of bowed notes is changed most with the left hand for any right-handed cellist. Still, the bilateral activity we find in dPMC and SMA resonates with the fact that playing a cello is bimanual, even if one hand (left) is more responsible for pitch change than the other (right).

4.3 Evidence for Embodied Cognition

The differences between the discriminating neural activity of experts versus that of novices raises the issue of whether the cellists' discrimination is linked to an embodied cognition of the audio stimulus. The primary thesis of embodied cognition is that all aspects of cognition (such as thought, perception, reasoning, etc.) are based on the fact that the brain is situated in a body that interacts with an external environment (Borghi & Cimatti, 2010; Clark, 1997; Liberman & Mattingly, 1985) through sensorimotor systems. The opportunity of studying expert cellists alongside relatively musically naïve novices allows us to provide evidence for or against embodied cognition because of the experts' acute sensorimotor-cognitive systems, developed and maintained over years of both musical and instrumental training from an early age. Furthermore, we can directly probe this acute sensorimotor-cognitive system using an auditory stimulus (the J.S. Bach piece) that has served a developmental and maintenance function over the years, since it has a perennial place in the repertoire of cellists from beginner students to expert professionals, such as those used in our study.

There has been much work on embodied cognition in psychology. For instance, Olmstead et al (Olmstead, Viswanathan, Aicher, & Fowler, 2009) have shown differences in behavioral response to sentence comprehension (specifically, plausible vs. implausible action-related sentences). Although measured without electrophysiological data, sentence comprehension paradigms provide a relevant analogy to our experiment using key-changes because an anomalous event appears in a temporal sequence (e.g., a grammatical error or non-sequitur) and must be recognized as such, given background knowledge of the stimuli's evolution (e.g., semantic sense) and any relevant preceding stimuli (e.g., the preceding words). As discussed earlier, such a semantic-related task would likely invoke later components, such as the N400 or N600, as we find in traditional ERP analysis (see Appendix).

In the case of our experiment, the background knowledge base is the primary difference between our subject classes. The experts have years of formal auditory and motor training from an early age to play the cello. This extensive training has likely led to a long-term memory of the J.S. Bach piece on many levels of auditory and motor nuance that is called upon when prodded with an auditory stimulus, such as a recording of the piece, unfolding a sequence of expectations with each uptake of audio sampling in what they hear. We interfere with this process each time the key is changed, thereby prodding this acute system even further. The novices, on the other hand, simply do not have the training or continued audio-motor maintenance regimen in place to comprehend the musical stimulus at a comparable level of nuance. Having ruled out other possible sources of differentiation, this difference likely contributes to the different class-wise neural and behavioral responses as the two groups identify AMEs.

Claims of proof for embodied cognition have also been made in neuroscience, though not yet with the same prevalence as in other disciplines. Recent work on mirror neuron systems (Lotto, Hickok, & Holt, 2009) has opened the door to such analysis. Stemming from such

work and overlapping in stimulus type with Olmstead, Tettamanti et al. (Tettamanti et al., 2005) have shown fMRI evidence for fronto-parietal motor circuit activation in response to reading action-related sentences. By showing a connection to motor circuitry from reading action-related sentences, the authors claim a link to embodied cognition using differential BOLD response. The parallels between our study in music cognition and this work by Tettamanti are similar to what they were for the study of Olmstead et al. (Olmstead et al., 2009), regarding the different knowledge bases of the two populations.

Our study supports the theory of embodied cognition by going beyond these earlier studies and examining the rapid neural dynamics (all within 1s of the stimulus) that identify one group as the expert class. Our technique of manipulating the cellists' profound expertise via the introduction of an anomalous musical event (i.e., a key-change) allows us to perturb and then to observe an expert's cognitive process in action. Rather than primarily basing our conclusions only on *post hoc* behavioral metrics (as did Olmstead et al. (Olmstead et al., 2009) and others), or utilizing the slow and TR-constrained BOLD response among subjects with potentially broad ranges of expertise (as did Tettamanti et al. (Tettamanti et al., 2005) and others), we have focused on the fleeting neural markers best measured with EEG (and possibly MEG) that would likely precede any behavioral response. Furthermore, we have done so in a population of expert subjects whose cognitive systems are highly specialized to perform the chosen stimulus. Although earlier studies have looked at these markers in the aggregate and/or over long periods of time in comparison to the underlying neural dynamics, our approach employs signal detection theory applied to EEG to more precisely localize these task-relevant dynamics both in time and in sensor/voxel space.

The discriminating neuronal sources of experts originating in bilateralized sensorimotor systems – with particular strength in the right lateralized motor areas controlling left arm and hand movement – provide evidence for a different cognitive process occurring in cellists vs. that occurring in novices, one that is dependent on their experience playing the cello. Furthermore, this activity in the experts' brains happens at the post-stimulus window when the key-change activity is maximally discriminating from corresponding control activity (i.e., maximum A_2) and at no other time pre- or post-stimulus time. While other studies have found activation over different parts of the expert brain when listening to music (Koelsch, 2009; Koelsch & Siebel, 2005; Loui, Li, & Schlaug, 2011), the highly specialized motor and somatosensory response in these experts listening to a piece they have played for many years on an instrument they have likewise played for many years cannot be underestimated as a piece of supporting evidence for embodied cognition. Furthermore, contemporaneous to our study, other researchers have found evidence for embodied syntax processing using transcranial magnetic stimulation (Candidi, Sacheli, Mega, & Aglioti, 2012) and traditional ERP analysis (Sammler, Novembre, Koelsch, & Keller, 2013).

Of course, there are several important caveats to consider relative to evidence we have found for embodied cognition. The first is the relatively small sample size for the two groups (5 experts vs. 5 novices), constrained by the number of professional concert cellists we were able to recruit for our study. Second is that additional controls and AMEs could be used to provide even stronger evidence for embodied cognition. For example, adding a control condition that includes both pieces played on the experts' native musical instrument and on an alternative instrument. AMEs can also be varied to include those that can be produced via the experts' instrument or cannot (e.g. a buzz or hiss added to the piece at specific times).

Embodied cognition proponents point to embodiment as an advantage the neural system has in understanding its world. If embodiment enhances perception, then we should see in our experiment that professional cellists perform better than novices. We corroborate this expectation on behavioral response and in the accompanying neural activity. Not only do we

see high accuracy rates among experts but we also find they have a greater number of discriminating points, earlier maximum discrimination, a higher maximum value of A_z , and a higher mean value of the discriminating component at maximum A_z than corresponding values in novices. Therefore, we believe not just from the sensorimotor specialization seen in expert neuronal sources, but also from the behavioral and neural discriminating metrics used, that the difference in embodied cognition ability between expert and novice subjects is the driving factor behind their superior neural discrimination and behavioral performance. Finally, the generality of our techniques need not only apply to musical expertise. Our methods could be generalized to study other classes of expert subjects, and thus to investigate whether this specialized cognitive embodiment exists in other areas of human knowledge and expertise.

5. Conclusions

In summary, we have identified neural markers to differentiate experts from novices in a musical context. These markers reflect a somatosensory and motor response in experts that coincides with better behavioral performance than seen in novices. We have shown evidence that this response is even specific to the type of stimulus (a key-change) used via a right lateralized motor and frontal response. Furthermore, our experts' behavioral and neural responses support theories of embodied cognition, possibly implying that neural signatures of expertise exist in other domains than music.

Acknowledgments

This work was supported a grant from the Army Research Office (W911NF-11-1-0219) and in part by an appointment to the U.S. Army Research Laboratory Postdoctoral Fellowship Program administered by the Oak Ridge Associated Universities through a contract with the U.S. Army Research Laboratory.

References

- Baumann S, Koeneke S, Schmidt CF, Meyer M, Lutz K, Jancke L. A network for audio-motor coordination in skilled pianists and non-musicians. *Brain Res.* 2007; 1161:65–78. [PubMed: 17603027]
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological).* 1995; 57(1): 289–300.
- Bidelman GM, Krishnan A, Gandour JT. Enhanced brainstem encoding predicts musicians' perceptual advantages with pitch. *Eur J Neurosci.* 2011; 33(3):530–538. [PubMed: 21198980]
- Bledowski C, Prvulovic D, Goebel R, Zanella FE, Linden DE. Attentional systems in target and distractor processing: a combined ERP and fMRI study. *Neuroimage.* 2004; 22(2):530–540. [PubMed: 15193581]
- Bledowski C, Prvulovic D, Hoechstetter K, Scherg M, Wibral M, Goebel R, et al. Localizing P300 generators in visual target and distractor processing: a combined event-related potential and functional magnetic resonance imaging study. *J Neurosci.* 2004; 24(42):9353–9360. [PubMed: 15496671]
- Borghi AM, Cimatti F. Embodied cognition and beyond: acting and sensing the body. *Neuropsychologia.* 2010; 48(3):763–773. [PubMed: 19913041]
- Calvo-Merino B, Glaser DE, Grezes J, Passingham RE, Haggard P. Action observation and acquired motor skills: an FMRI study with expert dancers. *Cereb Cortex.* 2005; 15(8):1243–1249. [PubMed: 15616133]
- Candidi M, Satchell LM, Mega I, Aglioti SM. Somatotopic Mapping of Piano Fingering Errors in Sensorimotor Experts: TMS Studies in Pianists and Visually Trained Musically Naives. *Cereb Cortex.* 2012

- Chennu S, Bekinschtein TA. Arousal modulates auditory attention and awareness: insights from sleep, sedation, and disorders of consciousness. *Front Psychol.* 2012; 3:65. [PubMed: 22403565]
- Clark, A. *Being There: Putting Brain, Body and World Together Again.* Cambridge, MA: MIT Press; 1997.
- Cummings A, Ceponiene R, Koyama A, Saygin AP, Townsend J, Dick F. Auditory semantic networks for words and natural sounds. *Brain Res.* 2006; 1115(1):92–107. [PubMed: 16962567]
- Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods.* 2004; 134(1):9–21. [PubMed: 15102499]
- Duda, R. *a Pattern Classification.* New York: J Wiley; 2001.
- Giuliano R, Pfordresher PQ, et al. Native experience with a tone language enhances pitch discrimination and the timing of neural responses to pitch change. *Frontiers in Psychology.* 2011; 2:146. [PubMed: 21886629]
- Goldman RI, Wei CY, Philiastides MG, Gerson AD, Friedman D, Brown TR, et al. Single-trial discrimination for integrating simultaneous EEG and fMRI: identifying cortical areas contributing to trial-to-trial variability in the auditory oddball task. *Neuroimage.* 2009; 47(1):136–147. [PubMed: 19345734]
- Gonsalves B, Paller KA. Brain potentials associated with recollective processing of spoken words. *Mem Cognit.* 2000; 28(3):321–330.
- Haslinger B, Erhard P, Altenmuller E, Hennenlotter A, Schlaug G, Graf von Einsiedel H, et al. Reduced recruitment of motor association areas during bimanual coordination in concert pianists. *Hum Brain Mapp.* 2004; 22(3):206–215. [PubMed: 15195287]
- Hasson U, Nusbaum HC, Small SL. Brain networks subserving the extraction of sentence information and its encoding to memory. *Cereb Cortex.* 2007; 17(12):2899–2913. [PubMed: 17372276]
- Holmes AP, Blair RC, Watson JD, Ford I. Nonparametric analysis of statistic images from functional mapping experiments. *J Cereb Blood Flow Metab.* 1996; 16(1):7–22. [PubMed: 8530558]
- Jordan, MIAJ; R, A. Hierarchical mixtures of experts and the EM algorithm. *Neural Computation.* 1994; 6:181–214.
- Katahira K, Ablat D, Masuda S, Okanoya K. Feedback-based error monitoring processes during musical performance: an ERP study. *Neurosci Res.* 2008; 61(1):120–128. [PubMed: 18359117]
- Koelsch S. Music-syntactic processing and auditory memory: similarities and differences between ERAN and MMN. *Psychophysiology.* 2009; 46(1):179–190. [PubMed: 19055508]
- Koelsch S, Kasper E, Sammler D, Schulze K, Gunter T, Friederici AD. Music, language and meaning: brain signatures of semantic processing. *Nat Neurosci.* 2004; 7(3):302–307. [PubMed: 14983184]
- Koelsch S, Schmidt BH, Kansok J. Effects of musical expertise on the early right anterior negativity: an event-related brain potential study. *Psychophysiology.* 2002; 39(5):657–663. [PubMed: 12236333]
- Koelsch S, Schroger E, Tervaniemi M. Superior pre-attentive auditory processing in musicians. *Neuroreport.* 1999; 10(6):1309–1313. [PubMed: 10363945]
- Koelsch S, Siebel WA. Towards a neural basis of music perception. *Trends Cogn Sci.* 2005; 9(12): 578–584. [PubMed: 16271503]
- Leaver AM, Van Lare J, Zielinski B, Halpern AR, Rauschecker JP. Brain activation during anticipation of sound sequences. *J Neurosci.* 2009; 29(8):2477–2485. [PubMed: 19244522]
- Lieberman AM, Mattingly IG. The motor theory of speech perception revised. *Cognition.* 1985; 21(1): 1–36. [PubMed: 4075760]
- Lotto AJ, Hickok GS, Holt LL. Reflections on mirror neurons and speech perception. *Trends Cogn Sci.* 2009; 13(3):110–114. [PubMed: 19223222]
- Loui P, Grent-'t-Jong T, Torpey D, Woldorff M. Effects of attention on the neural processing of harmonic syntax in Western music. *Brain Res Cogn Brain Res.* 2005; 25(3):678–687. [PubMed: 16257518]
- Loui P, Li HC, Schlaug G. White matter integrity in right hemisphere predicts pitch-related grammar learning. *Neuroimage.* 2011; 55(2):500–507. [PubMed: 21168517]

- Maidhof C, Vavatzanidis N, Prinz W, Rieger M, Koelsch S. Processing expectancy violations during music performance and perception: an ERP study. *J Cogn Neurosci*. 2010; 22(10):2401–2413. [PubMed: 19702473]
- Meister I, Krings T, Foltys H, Boroojerdi B, Muller M, Topper R, et al. Effects of long-term practice and task complexity in musicians and nonmusicians performing simple and complex motor tasks: implications for cortical motor organization. *Hum Brain Mapp*. 2005; 25(3):345–352. [PubMed: 15852385]
- Nichols TE, Holmes AP. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum Brain Mapp*. 2002; 15(1):1–25. [PubMed: 11747097]
- Olmstead AJ, Viswanathan N, Aicher KA, Fowler CA. Sentence comprehension affects the dynamics of bimanual coordination: implications for embodied cognition. *Q J Exp Psychol (Hove)*. 2009; 62(12):2409–2417. [PubMed: 19396732]
- Parra L, Alvino C, Tang A, Pearlmutter B, Yeung N, Osman A, et al. Linear spatial integration for single-trial detection in encephalography. *Neuroimage*. 2002; 17(1):223–230. [PubMed: 12482079]
- Parra LC, Spence CD, Gerson AD, Sajda P. Recipes for the linear analysis of EEG. *Neuroimage*. 2005; 28(2):326–341. [PubMed: 16084117]
- Pascual-Marqui RD. Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Methods Find Exp Clin Pharmacol*. 2002; 24(Suppl D):5–12. [PubMed: 12575463]
- Pascual-Marqui RD, Esslen M, Kochi K, Lehmann D. Functional imaging with low-resolution brain electromagnetic tomography (LORETA): a review. *Methods Find Exp Clin Pharmacol*. 2002; 24(Suppl C):91–95. [PubMed: 12575492]
- Pascual-Marqui RD, Lehmann D, Koenig T, Kochi K, Merlo MC, Hell D, et al. Low resolution brain electromagnetic tomography (LORETA) functional imaging in acute, neuroleptic-naive, first-episode, productive schizophrenia. *Psychiatry Res*. 1999; 90(3):169–179. [PubMed: 10466736]
- Pfordresher P. Coordination of perception and action in music performance. *Advances in Cognitive Psychology*. 2006; 2(2–3)
- Pfordresher P, Brown S. Enhanced production and perception of musical pitch in tone language speakers. *Attention Perception Psychophysics*. 2009; 71:1385–1398.
- Pfordresher P, Kulpa JD. The Dynamics of Disruption From Altered Auditory Feedback: Further Evidence for a Dissociation of Sequencing and Timing. *Journal of Experimental Psychology: Human Perception and Performance*. 2011; 37(3):949–967. [PubMed: 21480745]
- Ruiz MH, Jabusch HC, Altenmuller E. Detecting wrong notes in advance: neuronal correlates of error monitoring in pianists. *Cereb Cortex*. 2009; 19(11):2625–2639. [PubMed: 19276327]
- Sammler D, Novembre G, Koelsch S, Keller PE. Syntax in a pianist's hand: ERP signatures of 'embodied' syntax processing in music. *Cortex*. 2013; 49(5):1325–1339. [PubMed: 22832238]
- Schmithorst VJ, Holland SK, Plante E. Cognitive modules utilized for narrative comprehension in children: a functional magnetic resonance imaging study. *Neuroimage*. 2006; 29(1):254–266. [PubMed: 16109491]
- Sevdalis V, Keller PE. Captured by motion: dance, action understanding, and social cognition. *Brain Cogn*. 2011; 77(2):231–236. [PubMed: 21880410]
- Tettamanti M, Buccino G, Saccuman MC, Gallese V, Danna M, Scifo P, et al. Listening to action-related sentences activates fronto-parietal motor circuits. *J Cogn Neurosci*. 2005; 17(2):273–281. [PubMed: 15811239]
- Virtue S, Haberman J, Clancy Z, Parrish T, Jung Beeman M. Neural activity of inferences during story comprehension. *Brain Res*. 2006; 1084(1):104–114. [PubMed: 16574079]
- Wronka E, Kaiser J, Coenen AM. Neural generators of the auditory evoked potential components P3a and P3b. *Acta Neurobiol Exp (Wars)*. 2012; 72(1):51–64. [PubMed: 22508084]
- Zatorre RJ. There's more to auditory cortex than meets the ear. *Hear Res*. 2007; 229(1–2):24–30. [PubMed: 17341441]
- Zatorre RJ, Chen JL, Penhune VB. When the brain plays music: auditory-motor interactions in music perception and production. *Nat Rev Neurosci*. 2007; 8(7):547–558. [PubMed: 17585307]

Appendix

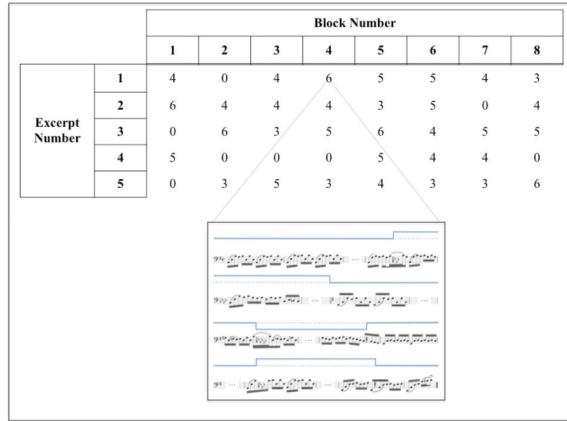


Figure A.1. Example of our randomized block design. Within each block, the ordering of excerpt playback was pseudorandomized. Blocks contained anywhere from zero to two control excerpts in which no key-changes occurred. The inset shows an example key-change excerpt for which six key-changes occurred. Although the shown inset begins and ends in the original recording key (G), this was not always the case, so as not to bias the detection simply being in the non-original recording key.

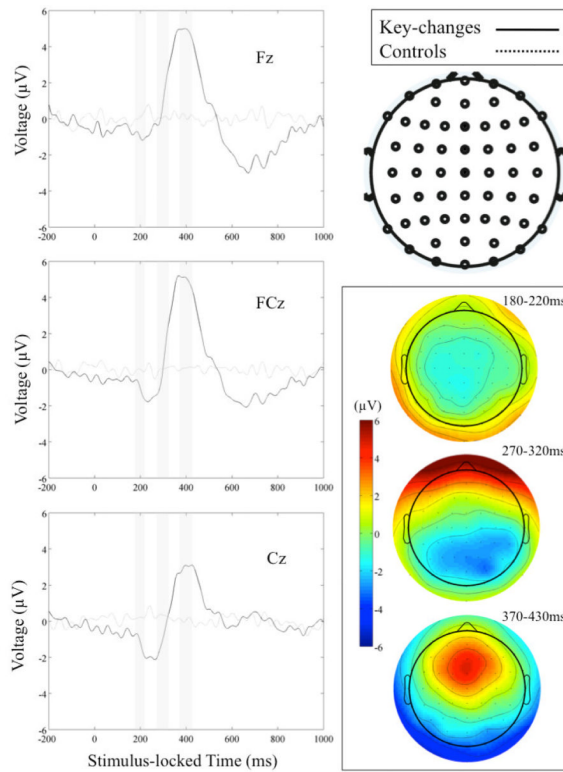


Figure A.2. Grand average ERPs of experts for key-change and control conditions. At selected electrodes (Fz, FCz, and Cz), we see an ERAN just after 200ms that is more pronounced in Cz relative to the frontal electrodes. The dominating component for each electrode is the P3, peaking around 400ms. Finally, the late negativity likely due to semantic

content processing (e.g., N400, N600, etc.) is most pronounced in frontal sites. Inset shows scalp ERPs at indicated time windows chosen *a priori* as in Maidhof et al. (Maidhof et al., 2010).

Table A.1

Parameters used in the sLORETA statistical analysis. With 10 samples per subject group, the maximum number of permutations for SnPM is 2^{10} . This represents the number of randomizations we used in our analysis. Our statistical analysis used the log of ratio of averages and an independent groups t-test.

sLORETA Parameters Name	Value
No normalization	TRUE
Independent groups, test A=B	TRUE
No baseline	TRUE
All tests for all Time-Frames/Frequencies	TRUE
Log of ratio of averages (similar to log of f-ratio)	TRUE
Number of randomizations	1024

Highlights

- We identify neural markers for musical expertise using single-trial analysis of the EEG
- We show that experts and novices have different spatio-temporal neural signatures of anomaly detection)
- We find evidence for an auditory-motor coupling in musical musical experts when detecting anomalies
- We provide evidence for the embodied cognition hypothesis

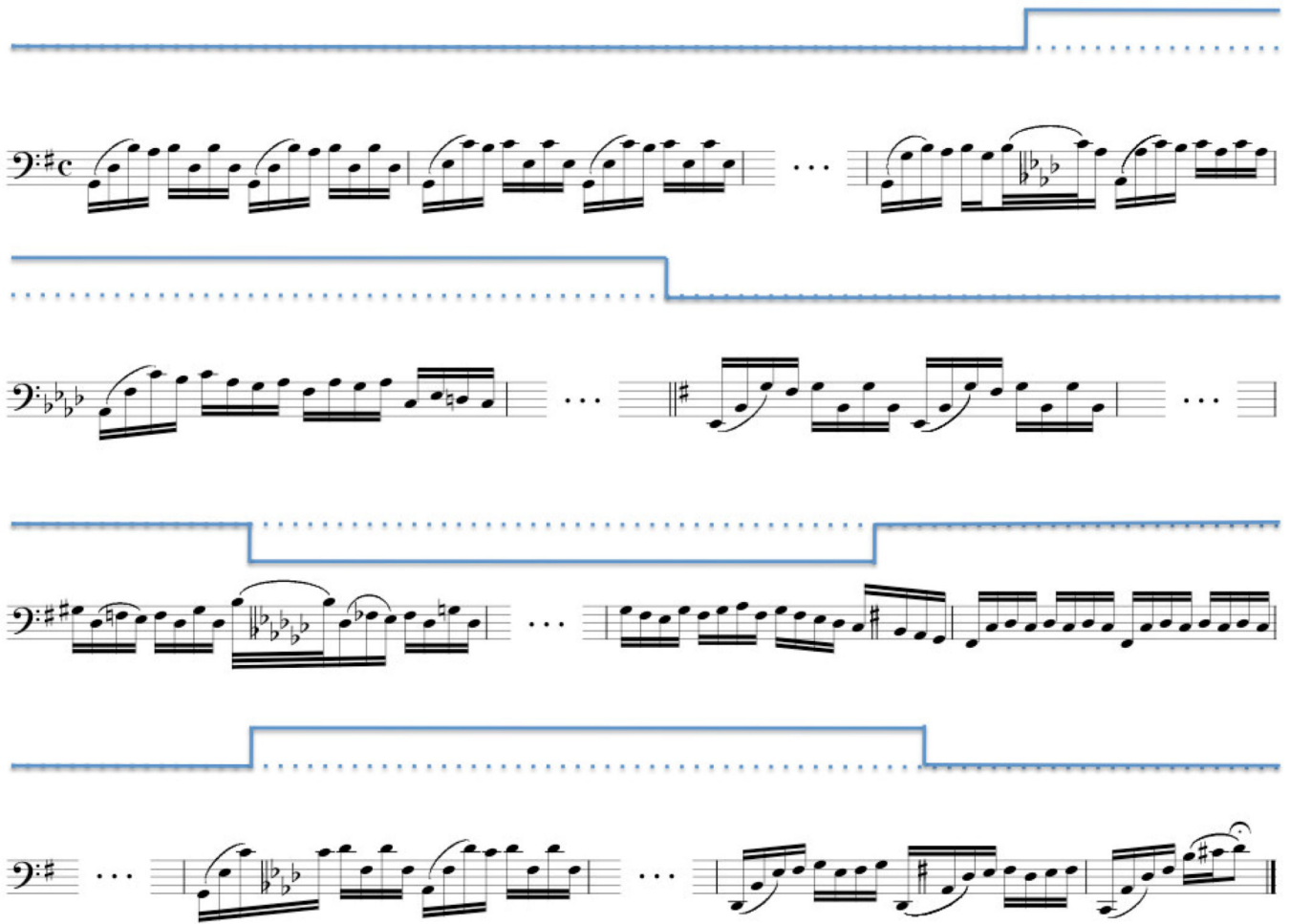


Figure 1. Musical notation and representative step-function of an alternation track showing the semi-tone key-changes imposed on Bach’s *Cello Suite No. 1, Prelude*. Above each staff, a step-function shows the key of the music with respect to the original key of G Major. The dotted-line represents the original key of G. Ellipses between bars represent music that remained in the preceding key. It is important to note from this figure that the key-changes were not limited to occur between notes or between musical phrases. Rather, they could occur during notes (e.g., the first and third) or between them, as well as between musical phrases. Although this example begins and ends in the original key, this was not true for the remaining alteration tracks. The total excerpt length is approximately 65 seconds.

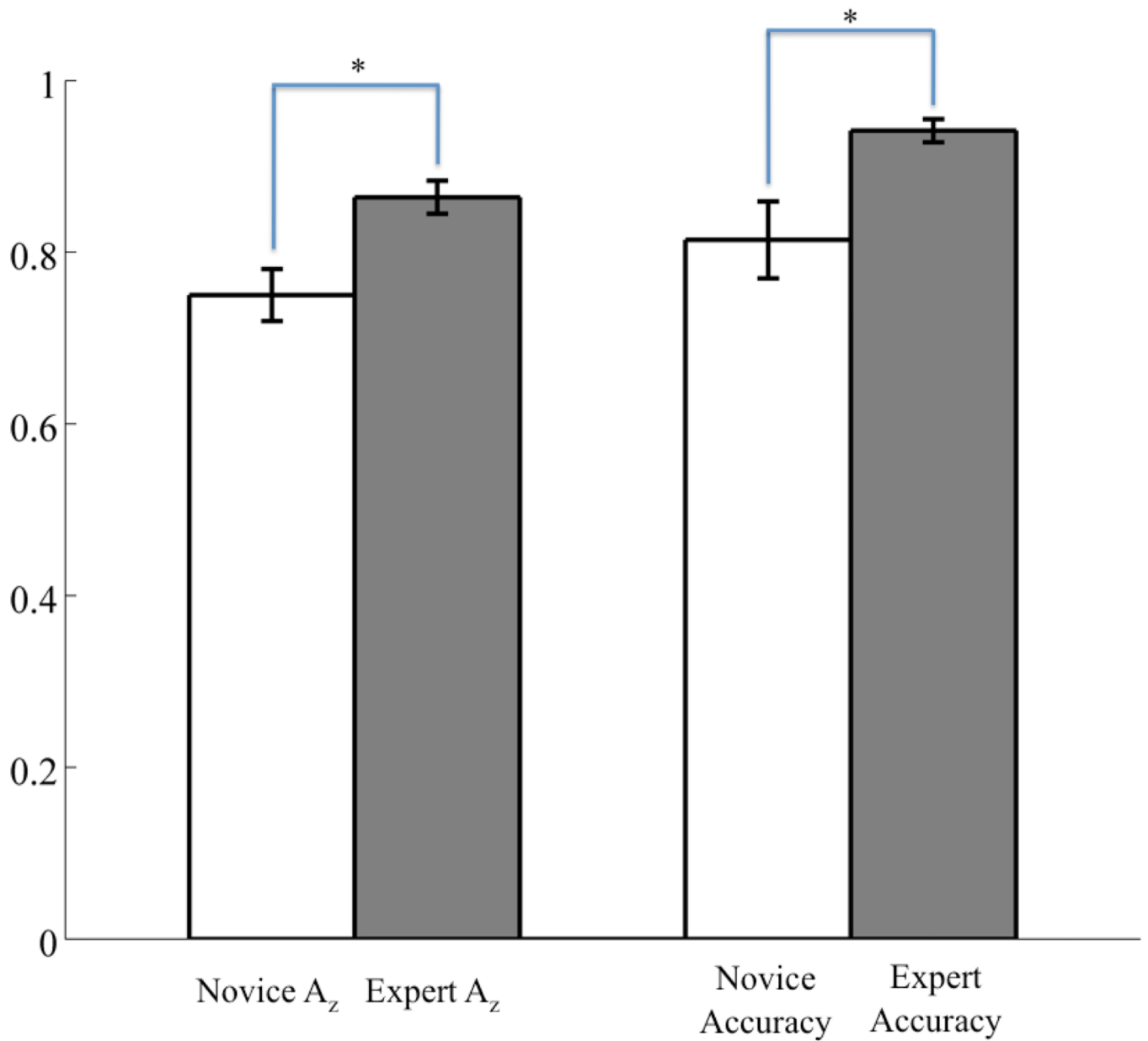


Figure 2. Comparison of neural discrimination to behavior between subject classes. White bars indicate novice subjects and grey bars indicate experts. Significant differences ($p < 0.02$, independent groups t-test) between subject classes is indicated with an asterisk (*). Note that accuracy is shown as a probability, rather than a percentage.

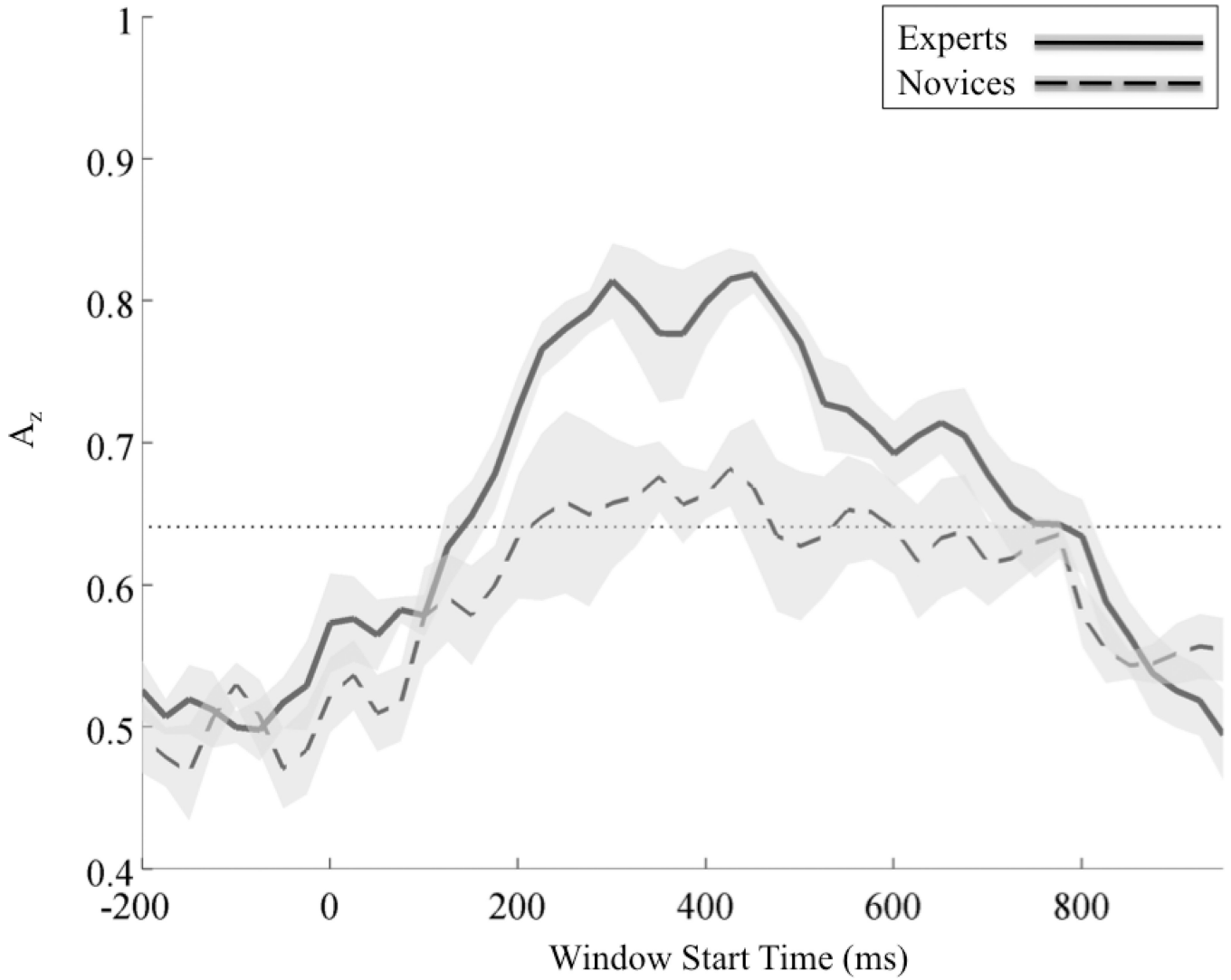


Figure 3. Stimulus-locked leave-one out discrimination for experts and novices. Each A_z curve shows the mean and standard error bands computed using leave-one-out discrimination for key-change vs. corresponding control events. The significance line (dotted) is corrected for multiple comparisons (line at $p=0.05$ Bonferroni corrected for 47 time window comparisons).

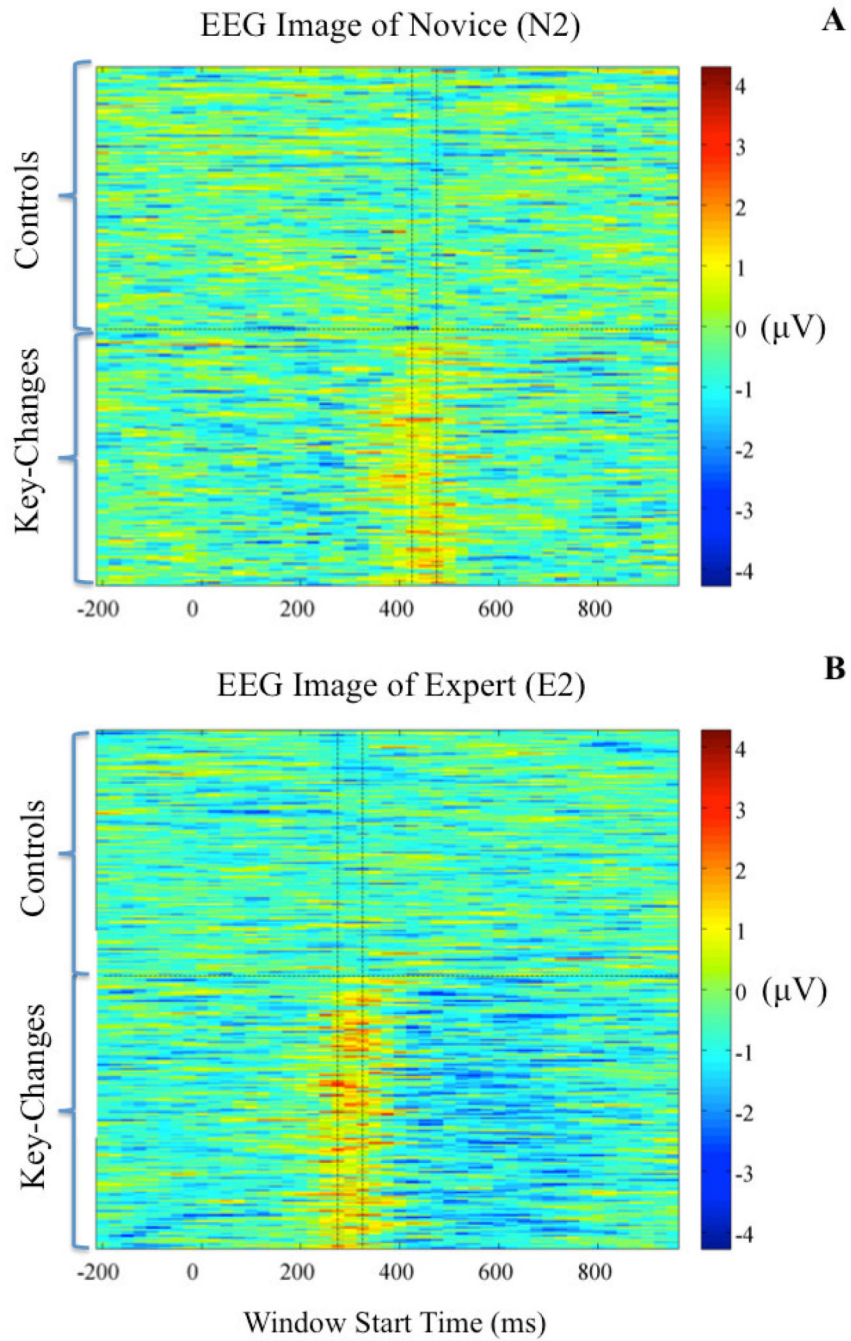


Figure 4. EEG Image of an age- and gender-matched novice (A) and an expert (B) showing the relative timing of the maximum discriminating EEG components. For both plots, color scale is in microVolts. The vertical dashed lines indicate the window of maximum discrimination between key-changes and controls. This subject pairing demonstrates the earlier timing of the maximum A_z seen in experts relative to novices, as well as the higher values of the discriminating component in experts that leads them to have a higher maximum A_z relative to novices.

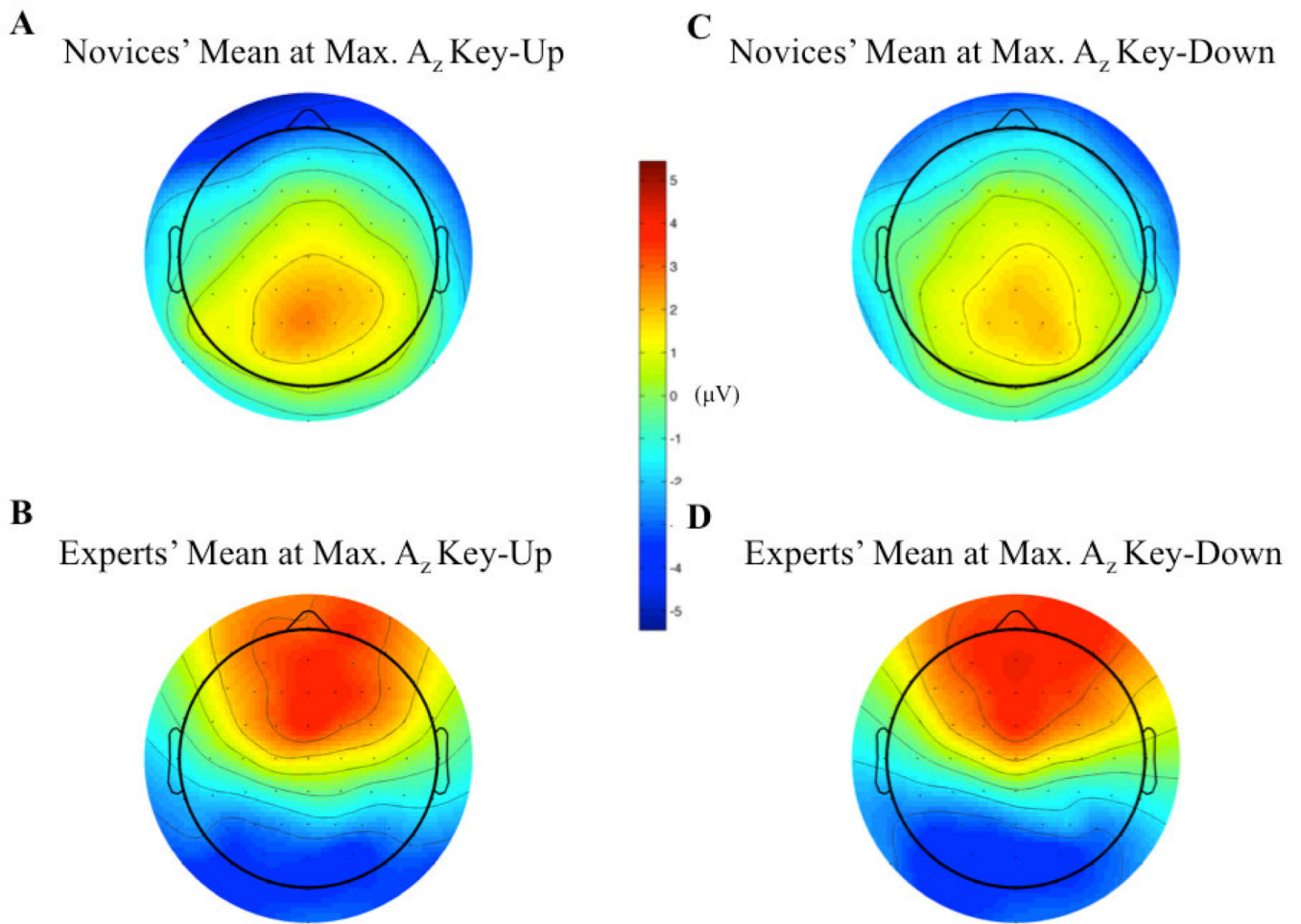


Figure 5. Grand average ERPs, for key-changes up and down, shown for expert and novice groups. There are no significant differences between key-changes up (A and B) and down (C and D) within groups ($p_{\text{experts}} > 0.90$ and $p_{\text{novices}} > 0.90$, Kolmogorov-Smirnov two-sample test). Novice's ERPs are consistent with a P300, while experts' are not.

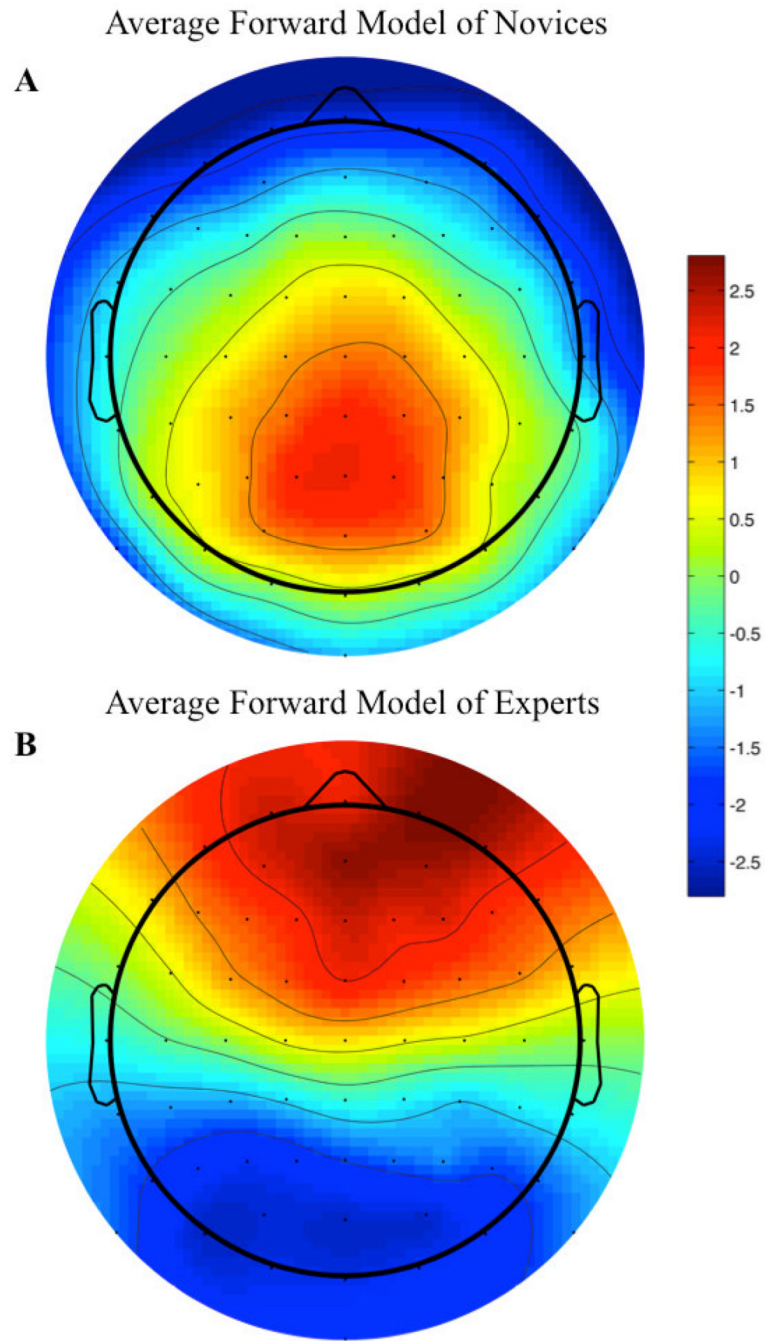


Figure 6. Average forward models for key-changes vs. controls. Forward models are first constructed for each subject using their specific temporal window of maximum Az. These subject specific forward models are then averaged to produce the group results shown above. For both plots, color scale is without units (see main text for discussion). The average expert forward models (B) exhibit strong frontal and frontal activation, while showing strong deactivation of occipital and occipito-parietal sensor sites for the target condition. Conversely, the average novice forward model (A) exhibits strong frontal deactivation, while showing strong occipital activation for the target condition.

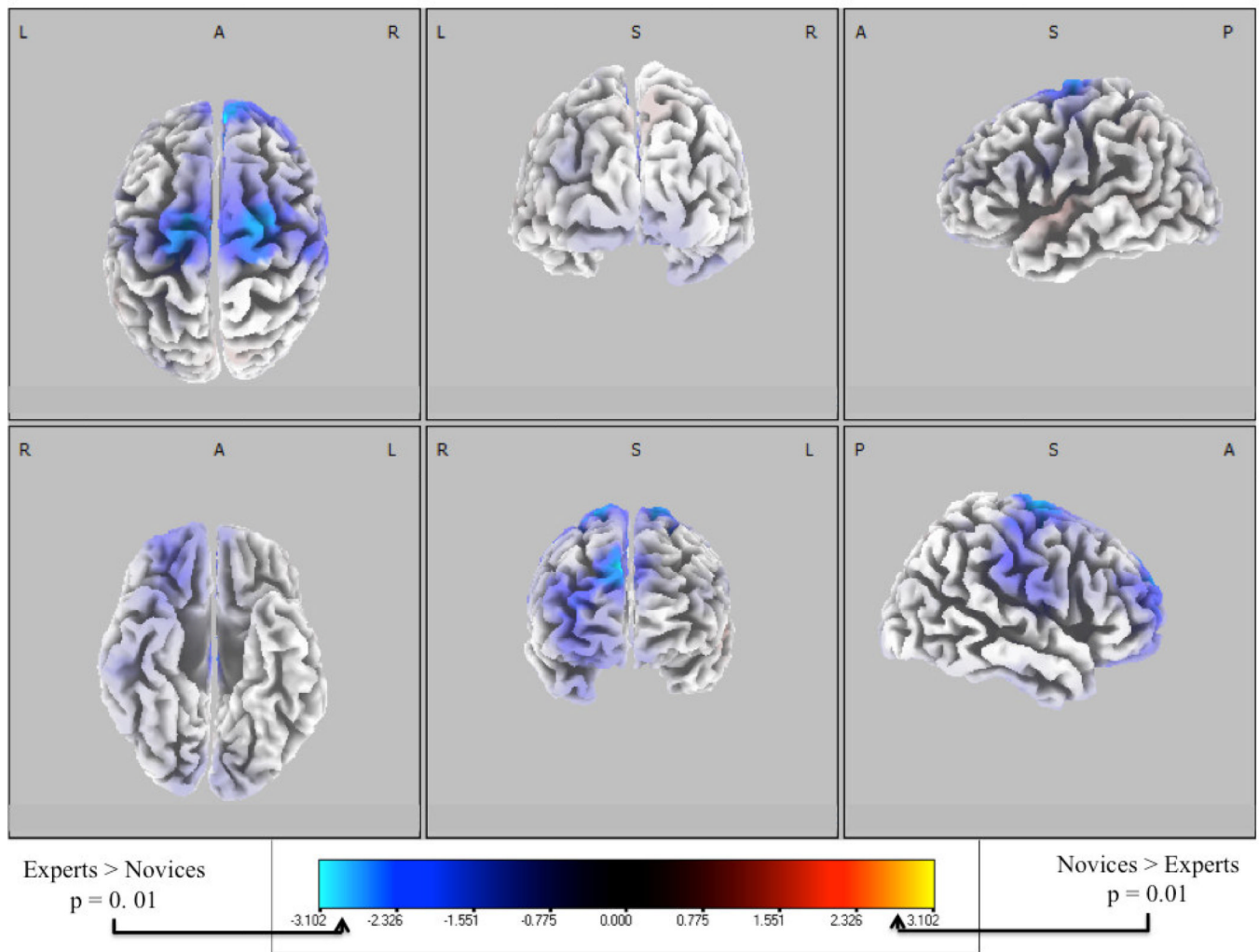


Figure 7. Six-views of neuronal current independent groups t-tests, with comparison between experts and novices at peak single-trial EEG discrimination. The t-distribution values of the log of the ratio of averages are shown for each voxel. One-tailed comparisons within each population class show neuronal current sources particularly strong for experts in BA 6 and 9 at the window of maximum Az. For key-change conditions, experts exhibited greater right-lateralized activation of neuronal current sources than novices ($p < 0.01$, independent groups t-test), especially in the right motor and frontal cortices. No common sources exist for novices ($p = 0.73$, independent groups t-test) at peak discrimination.

Table 1

Summary of ages, formal musical training, and cello performance experience in years of all subjects.

Subject Identifier	Novices					Experts				
	N1	N2	N3	N4	N5	E1	E2	E3	E4	E5
Years of Formal Musical Training	0	0	0	0	0	45	35	20	44	30
Years of Cello Performance	0	0	0	0	0	41	32	20	44	26
Age in years	27	41	28	28	30	51	39	28	53	35

Summary of EEG discrimination performance for individual subjects in each group (experts or novices). The values of maximum A_z and the post-key-change times to this maximum are given for all novices (N1-N5) and experts (E1-E5). Not only are experts faster to this maximum ($p < 0.02$, independent groups t-test), but their values of maximum A_z are also higher than those of novices ($p < 0.01$, independent groups t-test).

Table 2

Subject Identifier	N1	N2	N3	N4	N5	E1	E2	E3	E4	E5
Max. A_z	0.82	0.79	0.72	0.76	0.65	0.89	0.86	0.81	0.84	0.92
Time to Max. A_z in ms	475	450	375	675	775	425	300	250	300	350

Table 3

MNI coordinates, Brodmann areas and cortical structures showing greater neuronal source activity among experts than novices at peak discrimination between key-changes and controls. After correcting for multiple comparisons using SnPM, the 15 points shown here are the only voxels at which experts' neuronal sources are greater than those of novices ($p < 0.01$, independent groups t-test).

X(MNI)	Y(MNI)	Z(MNI)	Voxel t-value	Brodman area	Structure
20	-10	60	-3.102	6	Sub-Gyral
-15	-10	55	-3.091	6	Medial Frontal Gyrus
5	55	35	-3.079	9	Superior Frontal Gyrus
10	55	35	-3.068	9	Superior Frontal Gyrus
-15	-15	55	-3.044	6	Medial Frontal Gyrus
20	-5	60	-3.008	6	Sub-Gyral
5	50	35	-2.999	9	Superior Frontal Gyrus
20	-5	65	-2.969	6	Middle Frontal Gyrus
25	-5	60	-2.968	6	Sub-Gyral
20	-10	65	-2.963	6	Middle Frontal Gyrus
-15	-15	60	-2.960	6	Medial Frontal Gyrus
25	-5	65	-2.941	6	Middle Frontal Gyrus
-15	-10	65	-2.939	6	Middle Frontal Gyrus
-10	-10	55	-2.933	6	Medial Frontal Gyrus
-20	-10	60	-2.902	6	Sub-Gyral