



Contents lists available at SciVerse ScienceDirect

# Brain and Cognition

journal homepage: [www.elsevier.com/locate/b&c](http://www.elsevier.com/locate/b&c)



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

Jason Sherwin<sup>a,b,\*</sup>, Paul Sajda<sup>a</sup>

<sup>a</sup> Department of Biomedical Engineering, Columbia University, New York, NY 10027, USA

<sup>b</sup> Human Research and Engineering Directorate, U.S. Army Research Laboratory, Aberdeen, MD 21001, USA

### ARTICLE INFO

**Article history:**  
Accepted 12 July 2013  
Available online xxx

**Keywords:**  
Electroencephalography (EEG)  
Expertise  
Single-trial analysis  
Pattern recognition  
Perceptual decision-making

### 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 an 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.

© 2013 Published by Elsevier Inc.

### 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,

\* Corresponding author at: Department of Biomedical Engineering, Columbia University, New York, NY 10027, USA.  
E-mail address: [jason.sherwin@columbia.edu](mailto:jason.sherwin@columbia.edu) (J. Sherwin).

80 2002; Koelsch, Schroger, & Tervaniemi, 1999). Pfordresher and others  
81 have examined the action-related processes of music production  
82 using behavioral measures (Pfordresher, 2006; Pfordresher &  
83 Kulpa, 2011). In fact, music perception experiments primarily focus  
84 on pitch, along with melody and harmony discrimination, looking  
85 at the neural markers (usually ERPs) for deviant tones, notes or  
86 chords and the abilities of subjects to recognize syntactically inaccu-  
87 rate musical sequences (Bidelman, Krishnan, & Gandour, 2011;  
88 Koelsch, 2009; Koelsch & Siebel, 2005; Koelsch et al., 2004; Loui,  
89 Grent-t-Jong, Torpey, & Woldorff, 2005; Maidhof, Vavatzanidis,  
90 Prinz, Rieger, & Koelsch, 2010).

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

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

129 In music, an audience can often identify when an expected se-  
130 quence of events does not occur, i.e., when a 'mistake' occurs. This  
131 is the role that a forced key modulation serves as the chosen AME.  
132 In this regard, it can bear a strong resemblance to an oddball-like  
133 paradigm and in such studies the pool of available subjects with  
134 the required level of expertise to perform the task is quite high.  
135 For example, the ability to perceive the difference between tones  
136 of vastly different frequency relies only on normal hearing amid di-  
137 rected attention (Wronka, Kaiser, & Coenen, 2012) (Chennu &  
138 Bekinschtein, 2012). But native speakers of a tonal language  
139 (experts) can excel at the task in comparison to non-speakers (nov-  
140 ices) when the tones are closer in frequency (Giuliano, Pfordresher,  
141 et al., 2011; Pfordresher & Brown, 2009). Similarly in music, there  
142 is a class of subjects (musical experts) for whom oddball-like  
143 stimuli embedded within a particular musical stimulus will evoke  
144 a different, and perhaps stronger, neural response than it will in  
145 another class of subjects (musical novices). Earlier studies cited

above have begun to elucidate this using ERP analysis, but ques-  
tions still remain regarding the specific roles of action-related pro-  
cesses 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 in-  
structed 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 ex-  
perts (3 females, 2 males) and five were novices (2 females, 3  
males). The size of our expert population was limited by the num-  
ber of concert-level cellists we were able to recruit for the study.  
Despite this limited number of subjects, we found statistically sig-  
nificant 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  $32.6 \pm 10.0$  years of cello performance  
training and had a mean age of  $41.2 \pm 10.6$  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 per-  
forming. The novices had a mean age of  $30.8 \pm 5.8$  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 mu-  
sic 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 ob-  
tained 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 s 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 ex-  
cerpts altered to contain 3–6 anomalous key-change events. There  
were 140 total key-change events in the experiment.

Anomalous key-shifts were added to the original 44.1 kHz .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  
44.1 kHz .wav file. Although the pitch of the .wav file is raised or  
lowered with this algorithm, the effect is a complete change of mus-  
ical 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 pre-  
lude (originally in G), the recorded key was altered no higher than

**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



**Fig. 1.** Musical notation and representative step-function of an alternation track showing the semitone key-changes imposed on Bach's *Cello Suite No. 1, Prelude*. Above each stave, 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 s.

207 G-sharp (G#) and no lower than G-flat (Gb). This was done to avoid  
 208 overt distortion of the original recording, thereby making a key-  
 209 change obvious from non-harmonic considerations (e.g., timbre).  
 210 A schematic representation of a key-change trial beginning and  
 211 ending in the original recording key is shown in Fig. 1. In the figure,  
 212 a blue<sup>1</sup> step-function above the musical notation provides a schematic  
 213 of the key for those not fluent in musical notation.

214 We balanced the direction of key-changes using a total of 69  
 215 key-changes up and 71 key-changes down. Considering each key-  
 216 change as a stimulus, the inter-stimulus-interval (ISI) was at least  
 217 6 s. To remove the potential for bias, especially among our experts,  
 218 some trials began one semitone down or up from the original key  
 219 to ensure that the task was specific to relative key-change, rather  
 220 than deviation from absolute key.

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

226 A Dell Precision 530 Workstation was used to present the audio  
 227 stimuli with E-Prime 2.0 (Sharpsburg, PA) and a stereo audio card.  
 228 The subjects sat in an RF-shielded room between two Harmon &  
 229 Kardon computer speakers (HK695-01, Northridge, CA) connected  
 230 to the Dell. Each subject was allowed to adjust the volume so that

231 they could comfortably perform the task (playback volume did not  
 232 exceed 80 dB).

233 Subjects performed a simple detection task in which they were  
 234 asked to respond covertly by counting the number of AMEs they  
 235 heard. Counting was used rather than an overt behavioral re-  
 236 sponse, such as a button-press, to minimize motor confounds in  
 237 the EEG. Counts provided an estimate of task performance, thereby  
 238 ensuring that subjects attended to the task. Subjects were not  
 239 explicitly instructed to detect a key-change, but simply to pay  
 240 attention for anything out of the ordinary—i.e. an anomaly. They  
 241 were instructed not to move during the task and they were moni-  
 242 tored for any movement both visually and via analysis of motor-re-  
 243 lated artifacts in the EEG. Stimulus events were passed to the EEG  
 244 recording through a TTL pulse in the event channel. In *post hoc*  
 245 analysis, stimulus events were added to the EEG of control tracks  
 246 at the times when they had occurred in the key-change tracks.

### 2.3. Data acquisition

247  
 248 EEG data was acquired in an electrostatically shielded room  
 249 (ETS-Lindgren, Glendale Heights, IL, USA) using a BioSemi Active  
 250 Two AD Box ADC-12 (BioSemi, The Netherlands) amplifier from  
 251 64 active scalp electrodes. Data were sampled at 2048 Hz. A soft-  
 252 ware-based 0.5 Hz high pass filter was used to remove DC drifts  
 253 and 60 and 120 Hz (harmonic) notch filters were applied to mini-  
 254 mize line noise artifacts. These filters were designed to be linear-  
 255 phase to minimize distortions. Stimulus events – specifically,  
 256 key-changes – were recorded on separate channels.

<sup>1</sup> For interpretation of color in Figs. 1 and 7, the reader is referred to the web version of this article.

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

264 In epoching the data, the average baseline was removed from  
265 1000 ms pre-stimulus, i.e., 1000 ms before the AME or its corre-  
266 sponding control time in the epoch. After epoching into stimulus  
267 and time-matched control events, an automatic artifact epoch  
268 rejection algorithm in EEGLAB (Delorme & Makeig, 2004) was  
269 run to remove all epochs that exceeded a probability threshold of  
270 5 standard deviations from the average.

## 271 2.4. Data analysis

### 272 2.4.1. Behavioral accuracy from post-excerpt reporting

273 We calculated behavioral accuracy based on a post-excerpt  
274 reporting of how many AMEs the subjects counted. Subjects re-  
275 ported this number after each of the 40 listening excerpts. We cal-  
276 culated accuracy by noting the deviation from the actual number of  
277 key-changes in each excerpt. For instance, if excerpt  $i$  contained  $n_i$   
278 key-changes and a subject reported  $n_i - k_i$  or  $n_i + k_i$  key-changes  
279 then this constituted an error of  $k_i$ . However, if the subject reported  
280  $n_i$  then this constituted an error of zero,  $k_i = 0$ . To summarize the  
281 performance of each subject, we subtracted from 1.00 (perfect  
282 accuracy) the total number of errors, normalizing by the total num-  
283 ber of actual key-changes that occurred (Eq. (1)). Despite the pos-  
284 sibility of accuracy being less than zero, this case would only be in  
285 the event of extremely poor behavioral performance and did not  
286 occur in our experiment.

287 Behavioral accuracy calculation

$$288 \text{Accuracy} = 1 - \frac{\sum_{i=1}^{40} k_i}{\sum_{i=1}^{40} n_i} \quad (1)$$

### 291 2.4.2. Single-trial analysis of the EEG

292 We performed a single-trial analysis of the filtered, epoched and  
293 artifact-removed EEG to discriminate between anomalous key-  
294 changes in either direction, regardless of starting and ending key,  
295 and their corresponding control epochs. With the lack of an overt  
296 response to the key-change, we necessarily included both hits and  
297 misses in these epochs, thereby making the discrimination  
298 challenging if the subject does not demonstrate sufficient behav-  
299 ioral performance on the task. Logistic regression was used to find  
300 an optimal projection in the EEG sensor space for discriminating  
301 between these two conditions over each sub-window of the entire  
302 epoch (Parra et al., 2002, 2005). Specifically, we defined a training  
303 window starting at either a pre-stimulus or post-stimulus onset  
304 time  $\tau$ , with a duration of  $\delta$ , and used logistic regression to esti-  
305 mate a spatial weighting vector  $\mathbf{w}_{\tau,\delta}$  which maximally discrimi-  
306 nates between sensor array signals  $X$  for each condition (e.g.,  
307 key-changes vs. controls). For our experiments, the duration of  
308 the training window ( $\delta$ ) was 50 ms and the window onset time  
309 ( $\tau$ ) was varied across time  $\mathbf{w}_{\tau,\delta}$  ms in 25 ms steps (50% overlap),  
310 thereby covering  $[-200, 1000]$ ms. This training window size and  
311 overlap has been successfully used in other implementations of  
312 this technique (Parra et al., 2002, 2005), as it allows a suitable bal-  
313 ance between local and global temporal EEG dynamics. We used  
314 the re-weighted least squares algorithm to learn the optimal dis-  
315 criminating spatial weighting vector  $\mathbf{w}_{\tau,\delta}$  (Jordan & Jacobs, 1994).

316 Projection equation for component

$$317 \mathbf{y} = \mathbf{w}_{\tau,\delta} X \quad (2)$$

320 The result is a 'discriminating component'  $\mathbf{y}$  that is specific to  
321 activity correlated with each condition while minimizing activity  
322 correlated with both task conditions such as early audio process-  
323 ing. The term 'component' is used instead of 'source' to make it  
324 clear that this is a projection of all activity correlated with the  
325 underlying source. In Eq. (2),  $X$  is an  $N \times T$  matrix ( $N$  sensors and  
326  $T$  time samples).

327 Once solving for optimal discriminating spatial vectors in each  
328 window we can compute the electrical coupling coefficients (Eq.  
329 (3)).

330 Sensor projection onto discriminating component

$$331 \mathbf{a} = \frac{X\mathbf{y}}{\mathbf{y} \cdot \mathbf{y}} \quad (3)$$

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

338 We calculated the 'EEG image' by applying  $\mathbf{w}_{\tau,\delta}$  to the EEG data  
339 of each window ( $X(\tau)$ ),  $\tau \in [-200, 950]$ ms. Given a fixed value of  $\tau$ ,  
340 the result of this calculation provides a trial-by-trial visual repre-  
341 sentation of the window during which the discriminating compo-  
342 nent is at its highest value (see Section 3).

343 We quantified the performance of the linear discriminator by  
344 the area under the receiver operator characteristic (ROC) curve, re-  
345 ferred to as  $A_z$ , with a leave-one-out approach (Duda & et al., 2001).  
346 The ROC is a curve of false positive rate vs. true positive rate, there-  
347 fore greater area values under this curve indicate more accurate  
348 classification. We used the ROC  $A_z$  metric to characterize the dis-  
349 crimination performance between key-change and corresponding  
350 control epochs while sliding our 50 ms training window from start  
351 times of  $-200$  ms to  $950$  ms post-stimulus (i.e., varying  $\tau$ ). This  
352 epoch size provided substantial time both before and after the  
353 stimulus to observe any possible neural correlate of an anticipation  
354 of the AME.

355 We quantified the statistical significance of  $A_z$  in each window  
356 ( $\tau$ ) via a permutation-based relabeling procedure. In particular,  
357 we randomized the truth labels between control and key-change  
358 epochs and retrained the classifier. This was done 250 times for  
359 each subject in each of the forty-seven 50 ms windows, yielding  
360 11,500 permutations for each subject. On a group level (10 sub-  
361 jects), this yields 115,000 permutations. On a subject-level signifi-  
362 cance analysis, we utilized the false discovery rate (Benjamini &  
363 Hochberg, 1995) at  $p = 0.05$ , unless otherwise specified. On a  
364 group-level significance analysis, we utilized the Bonferroni cor-  
365 rection at  $p = 0.05$  (i.e.,  $p = 0.05/47 = 0.001$ ). For both levels, the  
366 number of permutations provided a suitable distribution to gauge  
367 statistical significance, regardless of the number of multiple com-  
368 parisons in epoch-time.

### 369 2.4.3. Traditional ERP analysis

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

384 **3. Results**

385 **3.1. Behavioral performance shows experts out-perform novices**

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

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

403 **3.2. Single-trial analysis reveals differences in neural activity between**  
404 **novices and experts**

405 Using the sliding window logistic regression classifier (see  
406 Methods), we found only post-stimulus windows of significant dis-  
407 crimination for each group of subjects. Fig. 3 shows each group's  
408 mean discrimination vs. epoch time. On average the leave-one-  
409 out (LOO) discrimination at each window is substantially greater  
410 for experts than it is for novices ( $p \ll 0.01$ , paired  $t$ -test). From vi-  
411 sual inspection we see that there are more discriminating windows  
412 for experts, as well as higher peak discrimination.

413 To further quantify the differences between groups, we can  
414 examine subject-level LOO results. Setting the false discovery rate  
415 (FDR) for each subject to  $p = 0.01$ , we find that experts have more  
416 significant discriminating windows than do novices ( $p < 0.02$ , inde-  
417 pendent groups  $t$ -test). Table 2 shows the values and times of max-  
418 imum discrimination ( $A_z$ ) for each subject. All discrimination  
419 values and corresponding times in this table are FDR corrected at

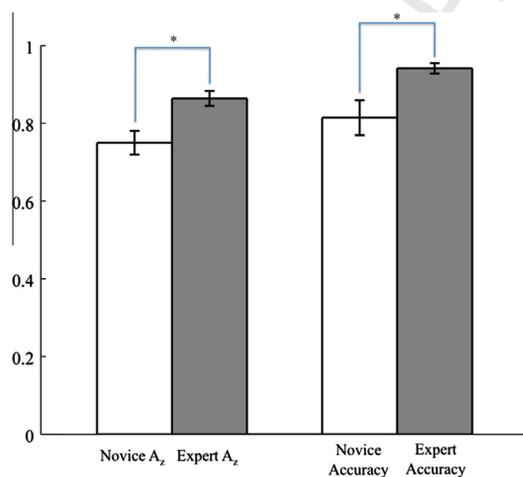


Fig. 2. Comparison of neural discrimination to behavior between subject classes. White bars indicate novice subjects and gray bars indicate experts. Significant differences ( $p < 0.02$ , independent groups  $t$ -test) between subject classes is indicated with an asterisk (\*). Note that accuracy is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

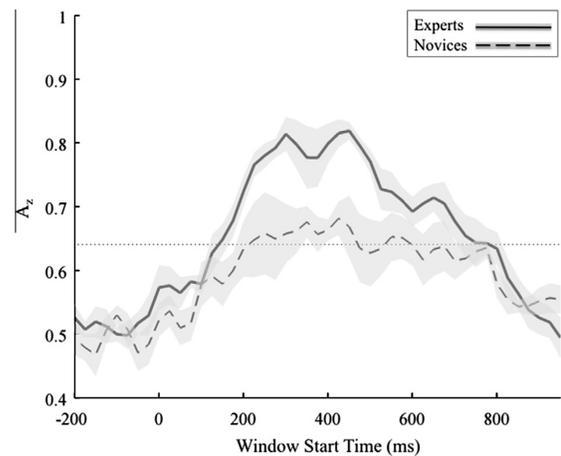


Fig. 3. Stimulus-locked leave-one out discrimination for experts and novices. Each  $A_z$  curve shows the mean and standard error bands computed using leave-one-out discrimination for key-change vs. corresponding control events. The significance line (dotted) is corrected for multiple comparisons (line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

420 a significance of  $p < 0.05$ . Experts ( $325 \pm 30$  ms) are faster to their  
421 maximum  $A_z$  than novices ( $550 \pm 75$  ms) ( $p < 0.02$ , independent  
422 groups  $t$ -test). Experts also exhibit higher values of maximum  $A_z$   
423 than do novices ( $p < 0.01$ , independent groups  $t$ -test), the latter  
424 of which can be seen in the behavioral accuracy differences in  
425 Fig. 2.

426 Finally, we utilized the discriminator's output to examine the  
427 single-trial variability of each subject to the key-change. We filtered  
428 each epoch's sensor data ( $X$ ) with the classifier ( $w_{\tau,\delta}$ ) that  
429 yielded the maximum value of discrimination ( $y$ ) between key-  
430 changes and controls (i.e., such that  $\tau$  is of maximum  $A_z$ ). This al-  
431 lowed us to examine the strength of the discriminating compo-  
432 nent, as well as its trial-to-trial variability. Generally, we found a  
433 common window in key-change trials across which each subject  
434 showed high values of discriminating activity compared both to  
435 neighboring windows and corresponding control trials' windows.  
436 We also found differences between experts and novices. For in-  
437 stance, Fig. 4 shows the discriminator outputs (mean of  $y$  in Eq.  
438 (2) across the window) of an age- and gender-matched novice  
439 and expert. From this figure, we see the demonstrated timing dif-  
440 ference of maximum  $A_z$  between novices (Fig. 4A) and experts  
441 (Fig. 4B). Furthermore, we see that the discriminating compo-  
442 nent for the expert during key-change trials is generally greater than  
443 that of the novice at the window of maximum discrimination. This  
444 observation also extends to the group level, where we find that the  
445 mean of the window-measured  $y_{\max A_z}$  during key-change trials is  
446 greater ( $p < 0.05$ , independent groups  $t$ -test) for experts  
447 ( $2.72 \pm 0.14 \mu V$ ) than it is for novices ( $2.22 \pm 0.23 \mu V$ ), indicating  
448 stronger discriminating activity among the experts at peak neural  
449 response to the key-change.

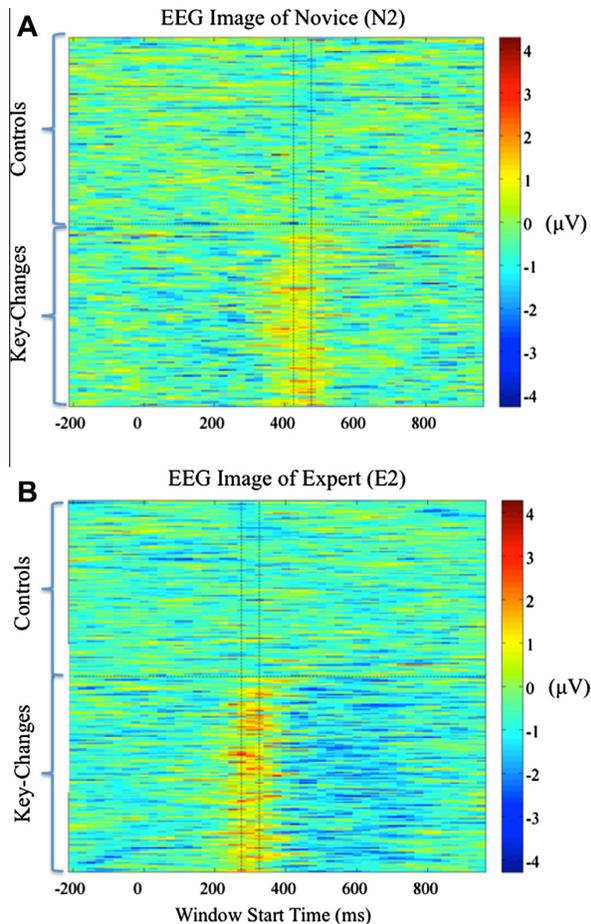
450 **3.3. Traditional ERP analysis shows group differences at peak**  
451 **discrimination**

452 We utilized our discriminator ( $w_{\tau,\delta}$ ) to examine the grand aver-  
453 age ERP of each group at subject-specific times of peak discrimina-  
454 tion. Since each time is actually the start time of a 50 ms window,  
455 we averaged across time within this window of peak  $A_z$ . We also  
456 considered the differences, if any, between up and down key-  
457 changes for either group. After epoch rejection, we found no signif-  
458 icant differences between up ( $64 \pm 1$ ) and down ( $65 \pm 1$ ) trials for

**Table 2**

Summary of EEG discrimination performance for individual subjects in each group (experts or novices). The values of maximum  $A_2$  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  $A_2$  are also higher than those of novices ( $p < 0.01$ , independent groups  $t$ -test).

Subject identifier	N1	N2	N3	N4	N5	E1	E2	E3	E4	E5
Max $A_2$	0.82	0.79	0.72	0.76	0.65	0.89	0.86	0.81	0.84	0.92
Time to max $A_2$ in ms	475	450	375	675	775	425	300	250	300	350



**Fig. 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  $A_2$  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  $A_2$  relative to novices. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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. Fig. 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, Hoehstetter, et al., 2004), whereas the experts' activity is frontal. Importantly, in comparison to previous work (e.g., Koelsch, 2009;

Koelsch et al., 1999, 2002), we do not find the ERAN to be the most discriminating ERP component for musical experts, as their peak discrimination occurs at  $325 \pm 30$  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**

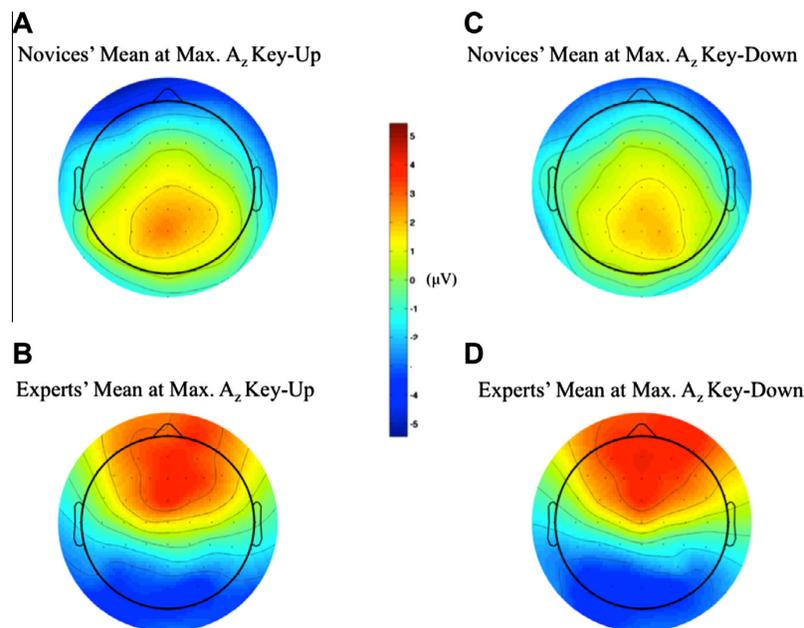
To more closely examine the differences between experts and novices in terms of the spatial distribution of their maximally discriminating components, we estimated the electrical coupling,  $\mathbf{a}$ , (i.e., the forward model) for each group. Fig. 6 shows these forward models for experts and novices computed using the components at each subject's maximum  $A_2$ . Specifically the forward model for each subject-class represents an average of the subject-specific forward models estimated using the window of maximum  $A_2$  (i.e. the times in Table 2 represent the  $\tau$ 's for estimating the components in Eq. (2) and the resulting forward models  $\mathbf{a}$  using Eq. (3)).

Clear from Fig. 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 Fig. 5A/B and C/D. The novice group (Fig. 6A) has a topology consistent with the posterior P300 (Bledowski, Prvulovic, Goebel, et al., 2004; Bledowski, Prvulovic, Hoehstetter, et al., 2004). As with the ERP results, this is consistent with what one might expect for a target vs. 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 Fig. 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  $A_2$  for up and down



**Fig. 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.

521 epochs, respectively. As for the ERP calculations, these are averaged  
522 across time in the 50 ms window of peak discrimination between  
523 key-changes and controls. We then compared the ten novice sLORETA  
524 fits with the ten corresponding expert fits. The sLORETA parameters  
525 used in the independent groups  $t$ -test can be found in Appendix A. We  
526 established significance using 1024 permutations and the SnPM procedure  
527 for voxel-space comparisons (Holmes, Blair, Watson, & Ford, 1996; Nichols & Holmes,  
528 2002; Pascual-Marqui, 2002; Pascual-Marqui et al., 1999). Fig. 7  
529 shows the  $t$ -distribution values of the log of the ratio of averages  
530 (similar to a one-way ANOVA,  $F(1, 19)$ ) for experts > novices (purple/blue)  
531 and novices > experts (orange/yellow) mapped to a six-view projection  
532 of the cortex. We find that experts have 15 voxels across BA 6 and BA 9  
533 with significantly greater activity than novices ( $p < 0.01$ , independent  
534 groups  $t$ -test), peaking at MNI (20, -10, 60). Interestingly, this voxel  
535 of peak activity has also been implicated in music imagery tasks in  
536 pianists (Baumann et al., 2007), though in that work bilateral  
537 activation was found in this part of the pre-motor dorsal cortex. For  
538 our results, not only is peak activity for experts in right motor cortex,  
539 but we also find more right than left lateralized activation when we  
540 consider SnPM-corrected voxels out to  $p < 0.05$  (31 left voxels vs.  
541 50 right voxels, 17 of which are in right frontal cortices). All 31  
542 left voxels are in the motor cortices, while right voxels are distributed  
543 between the right motor (33 voxels) and frontal cortices (17 voxels).  
544 Table 3 gives MNI coordinates for a subset of these voxels with their  
545  $t$ -distribution values (thresholded by  $p < 0.01$ ). There are no voxels  
546 showing significance for novice activity greater than experts ( $p = 0.73$ ).

549 We also tested the time-uniqueness of this response amongst  
550 experts by considering windows of non-maximum  $A_z$ . Such a test  
551 addresses possible concerns that the motor response seen in the  
552 experts (cellists) is only from them listening to their instrument of  
553 expertise being played (Zatorre et al., 2007), rather than being an  
554 additional neural correlate of the key-change detection event. To this  
555 end, we randomly selected a pre-stimulus window (i.e.,  $\tau \in [-1000,$   
556  $-50]$ ms) for each subject, rather than the window of maximum  $A_z$ ,  
557 and performed the same sLORETA statistical analysis

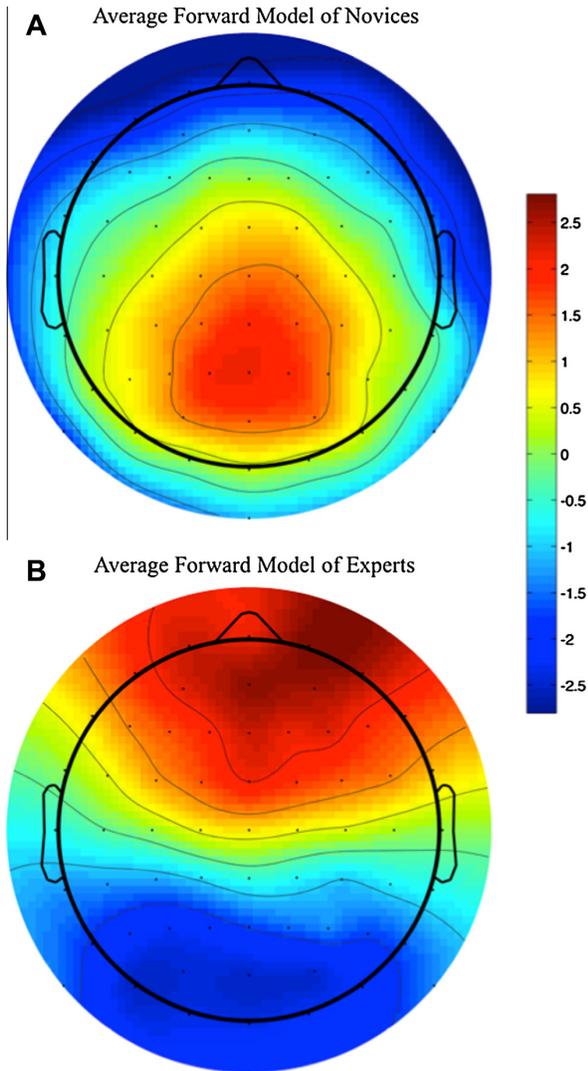
as before. We did the same calculation for a randomly chosen post-  
558 stimulus (i.e.,  $\tau \in [0, 950]$ ms) FDR-corrected insignificant and FDR-  
559 corrected significant window (non-maximum  $A_z$ ), respectively. For  
560 these three randomly selected window cases, we found no voxels  
561 having significantly greater activation for experts or novices ( $p > 0.05$ ,  
562 independent groups  $t$ -test, all three cases). 563

564 We also tested the robustness of our result among experts by  
565 averaging in the voxel space, rather than the sensor space, before  
566 statistical testing. Due to sLORETA's guarantee of zero-error in  
567 fitting the sensor distribution to the voxel space, and linearity of  
568 electromagnetic sources, we expected to duplicate our earlier findings.  
569 We transformed each epoch's time-averaged window of subject-  
570 specific maximum  $A_z$  into voxel space using sLORETA (ups and  
571 downs separately). We then averaged within subject all epochs  
572 for key-changes up and down, respectively. Performing the same  
573 SnPM and  $f$ -test, we found peak activity once again at MNI  
574 (20, -10, 60) and the same 15 voxels showing greater activity  
575 among experts than novices ( $p < 0.01$ , independent groups  $t$ -test). 576

577 Finally, the slight right lateralization of the neuronal current  
578 response among experts and the frontal source found from this analysis  
579 is consistent with both the ERP and forward model results, 579  
580 Figs. 5B/D and 6B, respectively.

#### 4. Discussion

580 In this paper, we have shown that both the timing and sources  
581 of discriminative neural markers are different between our expert  
582 and novice subjects in an oddball-like musical anomaly detection  
583 task. Without particular instruction to detect key-changes, all  
584 subjects detected the AMEs, with group differences manifesting  
585 themselves in behavioral performance, discriminating neural activity,  
586 traditional ERP analysis, scalp topology of discriminating  
587 component forward models and the distribution of neuronal sources.  
588 We now discuss these results in the context of relevant and related  
589 studies. 590



**Fig. 6.** Average forward models for key-changes vs. controls. Forward models are first constructed for each subject using their specific temporal window of maximum  $A_2$ . 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 occipitoparietal 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.

591 **4.1. Studying expertise in musicality**

592 Other studies have investigated the neural correlates of experts  
593 and novices with respect to musical stimuli. For instance, expert  
594 pianists were shown to have less fMRI activation than control sub-  
595 jects in pre-motor cortex during complex movement tasks at a pia-  
596 no keyboard, indicating a learning effect (Meister et al., 2005);  
597 pianists have been shown to have higher fMRI activation in motor  
598 areas when listening to musical stimuli than non-pianists (Bau-  
599 mann et al., 2007); EEG has been used to show neural signatures  
600 that precede when a trained pianist is about to hit an incorrect  
601 note (Ruiz et al., 2009). EEG also has been used to examine the role  
602 of auditory feedback in trained vs. untrained pianists, where it was

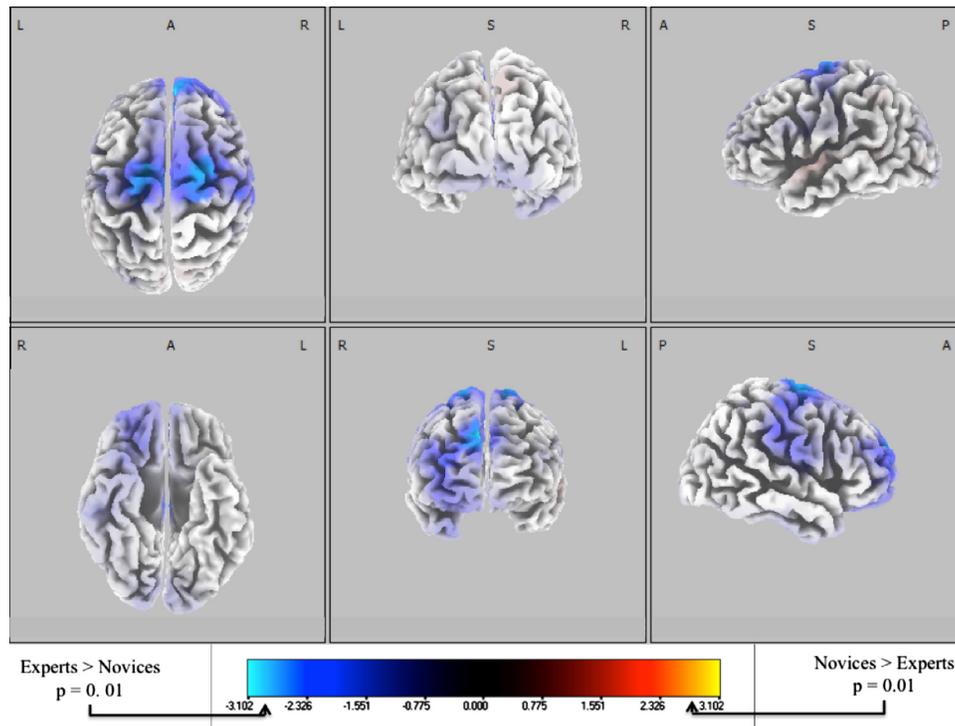
603 found that an N210 ERP was seen for experienced vs. less-experi-  
604 enced pianists following an alteration of the auditory feedback  
605 (Katahira, Abla, Masuda, & Okanoya, 2008). Many EEG studies also  
606 have investigated augmented pitch processing capabilities of ex-  
607 pert musicians (Koelsch et al., 1999, 2002).

608 While these previous studies provide hints of neural markers of  
609 musical expertise, their conclusions stem from the results of neural  
610 activity measured over periods of time that are long when com-  
611 pared to the underlying neural dynamics. For instance, the tempo-  
612 ral resolution of fMRI is typically constrained by the repetition  
613 time TR and the sluggish BOLD response. Traditional EEG studies  
614 on the other hand, while they reveal phenomena with high tempo-  
615 ral resolution, are mostly the result of averaging over many trials  
616 that unfold in time across many minutes, ignoring the variability  
617 of the activity across trials.

618 Nevertheless, we can directly compare our findings to the exist-  
619 ing EEG literature on music perception and expertise, at least  
620 where the experimental task is similar. Maidhof et al. (2010) em-  
621 ploy a task, similar to ours, in an attempt to separate the action-  
622 and perception-related processes in musical deviance, i.e., AME,  
623 detection. Since there is no action required by our subjects, the  
624 closest analog to our experimental design is from the 'perception  
625 condition' of this paper. Maidhof et al. however define an AME dif-  
626 ferently than we do in our experimental design. Specifically, in  
627 their paper, an AME is a single note that is flattened by one semi-  
628 tone from its diatonic scalar context, whereas our study's AME is a  
629 semitone key-change that does not return to the previous musical  
630 context after the event. Comparing our experts' EEG activity with  
631 those of Maidhof et al., we find comparable grand average ERPs  
632 in *a priori* ROIs (such as Fz, FCz and Cz) and scalp distributions in  
633 *a priori* post-AME time windows (e.g., 180–220 ms, 270–320 ms  
634 and 370–430 ms). These figures can be found in Appendix A for di-  
635 rect comparison to Fig. 3 of Maidhof et al. The primary differences  
636 between our ERPs and those of Maidhof et al. are that our N2 and  
637 P3 complexes are smaller in magnitude, though their timings are  
638 comparable. Our ERPs also exhibit a later negativity, most notable  
639 in Fz and FCz. These differences are likely due to the nature of the  
640 AME used in our experiment. In Maidhof et al., the AME lasts 104–  
641 217 ms but the harmonic context is quickly re-established. This is  
642 not the case in our experiment. Rather, we change key by forced,  
643 instead of diatonic, modulation. Consequently, the long-term har-  
644 monic expectations established from the larger musical context  
645 preceding the AME are not fulfilled. These contextual or seman-  
646 tic-related responses have been seen in music in the N400 and  
647 N500 (Koelsch, 2009; Koelsch et al., 2004; Loui et al., 2005) and  
648 in language in the N600 and P820 (Cummings et al., 2006; Gonsal-  
649 ves & Paller, 2000).

650 Other studies have specifically examined particular ERP compo-  
651 nents, such as the ERAN or MMN. Koelsch has looked at these com-  
652 ponents extensively in musical expertise (Koelsch, 2009; Koelsch  
653 et al., 2002). Although not a particular aim of our study, we can  
654 consider our results in the context of Koelsch's studies of the ERAN  
655 and MMN. We particularly focus on the ERAN because it is more  
656 dependent on music-syntactic regularities extant in long-term  
657 memory than the MMN, which depends on online establishment  
658 of regularities in auditory stimuli and is therefore not specific to  
659 music. We find concordant results with Koelsch (2009) and Loui  
660 et al. (2005) in that the ERAN of the experts is larger than the no-  
661 vices, especially in Fz and Cz (not shown).

662 Although we corroborate earlier results on musical expertise,  
663 we see some differences that are likely due to the differences in  
664 task. The most obvious difference between groups is in the P3  
665 amplitude, which is much larger for experts than novices in both  
666 electrode ROIs. Furthermore, the latency of the experts' P3 is smal-  
667 ler than that of the novices at Cz and is likely due to their higher  
668 sensitivity to the AME. We additionally find, through single-trial



**Fig. 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  $A_2$ . 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 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).

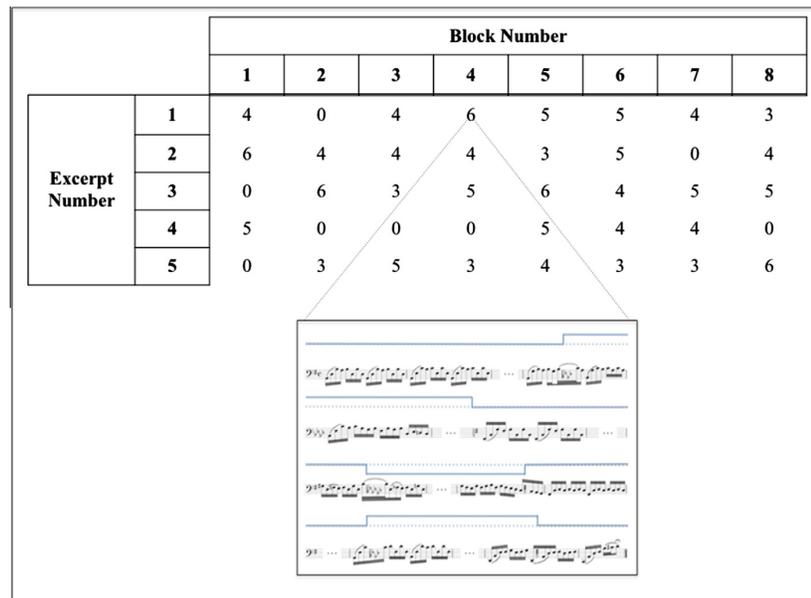
X (MNI)	Y (MNI)	Z (MNI)	Voxel <i>t</i> -value	Brodmann area	Structure
20	-10	60	-3.102	6	Sub-gyral
-15	-10	55	-3.091	6	Medial frontal gyrus
5	55	35	-3.079	9	Superior frontal gyrus
10	55	35	-3.068	9	Superior frontal gyrus
-15	-15	55	-3.044	6	Medial frontal gyrus
20	-5	60	-3.008	6	Sub-gyral
5	50	35	-2.999	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

669 analysis, that cellists peak in neural discrimination earlier than do  
670 the novices. Whereas previous ERP studies rely on electrode ROIs  
671 and time windows, we determine which time windows and elec-  
672 trodes are maximally discriminative between key-change and con-  
673 trol events. Finally, we found a larger late negativity (~600–  
674 700 ms) in the cellists than non-cellists (not shown). This late neg-  
675 ativity is not seen in Koelsch (2009), presumably because such late  
676 negativities are also associated with semantic content developed  
677 over longer periods of time (Cummings et al., 2006; Gonsalves &  
678 Paller, 2000; Koelsch, 2009; Koelsch et al., 2004; Loui et al.,

2005) than the stimuli they used. Compared to only 2.5 s of preced-  
ing 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 Appendix A). Our experimental design

679  
680  
681  
682  
683  
684  
685  
686  
687  
688



**Fig. A1.** 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.

689 focused on creating a stimulus salient to both groups. We found  
 690 that the expertise of the professional cellists allowed them to more  
 691 accurately and more quickly identify the anomalous events in a  
 692 sequential stream of musical stimuli. Therefore, a primary contribu-  
 693 tion of our paradigm and ensuing analysis is that it has focused  
 694 on the neural markers of AME detection that varied trial-to-trial  
 695 (see Fig. 4). We utilize this variability to identify discriminating  
 696 EEG components and the times of the maximum discriminating  
 697 components (Fig. 3 and Table 2), which ultimately inform our  
 698 ERP (Fig. 5), forward model (Fig. 6) and source localization  
 699 (Fig. 7) analyses. Consequently, while our results have been dem-  
 700 onstrated in the context of musicality, they can theoretically be ex-  
 701 tended to other domains of human interaction with the external  
 702 environment that depend on a tight coupling of sensorimotor  
 703 interaction to cognition (e.g., language comprehension and  
 704 production).

705 **4.2. Auditory–motor coupling in experts**

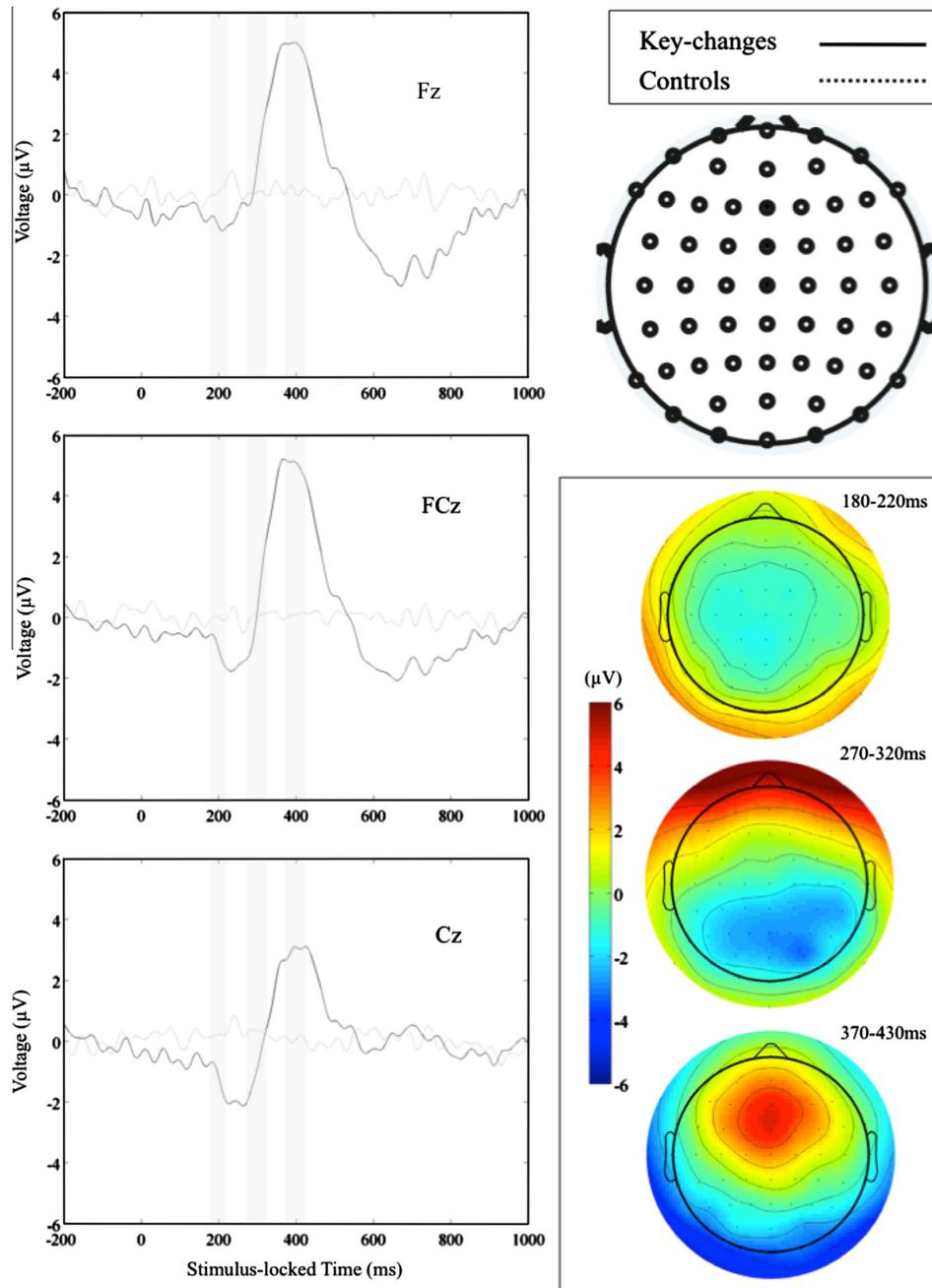
706 Our source localization results reveal strong activity amongst  
 707 experts in right lateralized frontal and bilateral motor, supplement-  
 708 ary motor and movement-related cortices (Fig. 7), though the  
 709 activity is more right than left lateralized. Significant values from  
 710 the comparison statistic ( $p < 0.01$ ) are found at locations impli-  
 711 cated in movement(s) and timing experiments involving the dorsal  
 712 pre-motor cortex (dPMC) and supplementary motor area (SMA).

713 In literature specific to music, there is a strong connection  
 714 emerging between audio and motor processing. Zatorre (2007)  
 715 and Zatorre et al. (2007) present a literature survey in which the  
 716 case is made for such a connection. They specifically focus on dor-  
 717 sal PMC (dPMC) as the location of mirror (or echo) neurons that  
 718 link auditory and motor systems. Our source localization is consis-  
 719 tent with this argument as we have also found an audio–motor  
 720 connection in the dPMC. They also claim that dPMC is involved  
 721 in the higher-order feature extraction of an auditory stimulus to  
 722 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 ex-  
 pert cellists, though it is possible that this difference is due to over-  
 familiarity with the piece. Leaver, Van Lare, Zielinski, Halpern, and  
 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 no-  
 vices was based more on a subjective notion of the prelude’s perme-  
 ation 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 associ-  
 ated 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 se-  
 quence manifests itself in an expert performer’s brain. 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



**Fig. A2.** Grand average ERPs of experts for key-change and control conditions. At selected electrodes (Fz, FCz, and Cz), we see an ERAN just after 200 ms that is more pronounced in Cz relative to the frontal electrodes. The dominating component for each electrode is the P3, peaking around 400 ms. 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. (2010).

759 non-maximally discriminating windows post-stimulus reinforces  
760 this role for SMA in key-change detection. The greater right later-  
761 alization of the activity in cellists further strengthens this claim,  
762 since the pitch of bowed notes is changed most with the left hand  
763 for any right-handed cellist. Still, the bilateral activity we find in  
764 dPMC and SMA resonates with the fact that playing a cello is  
765 bimanual, even if one hand (left) is more responsible for pitch  
766 change than the other (right).

#### 4.3. Evidence for embodied cognition

The differences between the discriminating neural activity of  
experts vs. 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, and reasoning)  
are based on the fact that the brain is situated in a body that inter-

**Table A1**

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.

Name	Value
<i>sLORETA parameters</i>	
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 <i>f</i> -ratio)	TRUE
Number of randomizations	1024

774 acts with an external environment (Borghi & Cimatti, 2010; Clark,  
775 1997; Liberman & Mattingly, 1985) through sensorimotor systems.  
776 The opportunity of studying expert cellists alongside relatively  
777 musically naïve novices allows us to provide evidence for or  
778 against embodied cognition because of the experts' acute sensori-  
779 motor-cognitive systems, developed and maintained over years of  
780 both musical and instrumental training from an early age. Further-  
781 more, we can directly probe this acute sensorimotor-cognitive sys-  
782 tem using an auditory stimulus (the J.S. Bach piece) that has served  
783 a developmental and maintenance function over the years, since it  
784 has a perennial place in the repertoire of cellists from beginner stu-  
785 dents to expert professionals, such as those used in our study.

786 There has been much work on embodied cognition in psychol-  
787 ogy. For instance, Olmstead, Viswanathan, Aicher, and Fowler  
788 (2009) have shown differences in behavioral response to sentence  
789 comprehension (specifically, plausible vs. implausible action-re-  
790 lated sentences). Although measured without electrophysiological  
791 data, sentence comprehension paradigms provide a relevant anal-  
792 ogy to our experiment using key-changes because an anomalous  
793 event appears in a temporal sequence (e.g., a grammatical error  
794 or non-sequitur) and must be recognized as such, given back-  
795 ground knowledge of the stimuli's evolution (e.g., semantic sense)  
796 and any relevant preceding stimuli (e.g., the preceding words). As  
797 discussed earlier, such a semantic-related task would likely invoke  
798 later components, such as the N400 or N600, as we find in tradi-  
799 tional ERP analysis (see Appendix A).

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

816 Claims of proof for embodied cognition have also been made in  
817 neuroscience, though not yet with the same prevalence as in other  
818 disciplines. Recent work on mirror neuron systems (Lotto, Hickok,  
819 & Holt, 2009) has opened the door to such analysis. Stemming  
820 from such work and overlapping in stimulus type with Olmstead, Tet-  
821 tamanti et al. (2005) have shown fMRI evidence for fronto-parietal  
822 motor circuit activation in response to reading action-related sen-  
823 tences. By showing a connection to motor circuitry from reading  
824 action-related sentences, the authors claim a link to embodied

825 cognition using differential BOLD response. The parallels between  
826 our study in music cognition and this work by Tettamanti are sim-  
827 ilar to what they were for the study of Olmstead et al. (2009),  
828 regarding the different knowledge bases of the two populations.

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

850 The discriminating neuronal sources of experts originating in  
851 bilateralized sensorimotor systems – with particular strength in  
852 the right lateralized motor areas controlling left arm and hand  
853 movement – provide evidence for a different cognitive process  
854 occurring in cellists vs. that occurring in novices, one that is depen-  
855 dent on their experience playing the cello. Furthermore, this activ-  
856 ity in the experts' brains happens at the post-stimulus window  
857 when the key-change activity is maximally discriminating from  
858 corresponding control activity (i.e., maximum  $A_2$ ) and at no other  
859 time pre- or post-stimulus time. While other studies have found  
860 activation over different parts of the expert brain when listening  
861 to music (Koelsch, 2009; Koelsch & Siebel, 2005; Loui, Li, & Schlaug,  
862 2011), the highly specialized motor and somatosensory response in  
863 these experts listening to a piece they have played for many years  
864 on an instrument they have likewise played for many years cannot  
865 be underestimated as a piece of supporting evidence for embodied  
866 cognition. Furthermore, contemporaneous to our study, other  
867 researchers have found evidence for embodied syntax processing  
868 using transcranial magnetic stimulation (Candidi, Sacheli, Mega,  
869 & Aglioti, 2012) and traditional ERP analysis (Sammler, Novembre,  
870 Koelsch, & Keller, 2013).

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

882 Embodied cognition proponents point to embodiment as an  
883 advantage the neural system has in understanding its world. If  
884 embodiment enhances perception, then we should see in our  
885 experiment that professional cellists perform better than novices.  
886 We corroborate this expectation on behavioral response and in  
887 the accompanying neural activity. Not only do we see high accu-  
888 racy rates among experts but we also find they have a greater num-  
889 ber of discriminating points, earlier maximum discrimination, a  
890 higher maximum value of  $A_2$ , and a higher mean value of the

891 discriminating component at maximum  $A_2$  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.

## 902 5. Conclusions

903 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.

## 913 Acknowledgments

914 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.

## 919 Appendix A

920 See Figs. A1 and A2 and Table A1.

## 921 References

- 922 Baumann, S., Koeneke, S., Schmidt, C. F., Meyer, M., Lutz, K., & Jancke, L. (2007). A network for audio-motor coordination in skilled pianists and non-musicians. *Brain Research*, 1161, 65–78.
- 923 Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)*, 57(1), 289–300.
- 924 Bidelman, G. M., Krishnan, A., & Gandour, J. T. (2011). Enhanced brainstem encoding predicts musicians' perceptual advantages with pitch. *European Journal of Neuroscience*, 33(3), 530–538.
- 925 Bledowski, C., Prvulovic, D., Goebel, R., Zanella, F. E., & Linden, D. E. (2004). Attentional systems in target and distractor processing: A combined ERP and fMRI study. *Neuroimage*, 22(2), 530–540.
- 926 Bledowski, C., Prvulovic, D., Hoehstetter, K., Scherg, M., Wibrall, M., Goebel, R., et al. (2004). Localizing P300 generators in visual target and distractor processing: A combined event-related potential and functional magnetic resonance imaging study. *Journal of Neuroscience*, 24(42), 9353–9360.
- 927 Borghi, A. M., & Cimatti, F. (2010). Embodied cognition and beyond: Acting and sensing the body. *Neuropsychologia*, 48(3), 763–773.
- 928 Calvo-Merino, B., Glaser, D. E., Grezes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: An fMRI study with expert dancers. *Cerebral Cortex*, 15(8), 1243–1249.
- 929 Candidi, M., Sachelì, L. M., Mega, I., & Aglioti, S. M. (2012). Somatotopic mapping of piano fingering errors in sensorimotor experts: TMS studies in pianists and visually trained musically naïves. *Cerebral Cortex*.
- 930 Chennu, S., & Bekinschtein, T. A. (2012). Arousal modulates auditory attention and awareness: Insights from sleep, sedation, and disorders of consciousness. *Frontiers in Psychology*, 3, 65.
- 931 Clark, A. (1997). *Being there: Putting brain, body and world together again*. Cambridge, MA: MIT Press.
- 932 Cummings, A., Cepenone, R., Koyama, A., Saygin, A. P., Townsend, J., & Dick, F. (2006). Auditory semantic networks for words and natural sounds. *Brain Research*, 1115(1), 92–107.
- 933 Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21.

- 934 Duda, R. et al. (2001). *Pattern classification*. New York: J Wiley.
- 935 Giuliano, R., Pfordresher, P. Q., et al. (2011). Native experience with a tone language enhances pitch discrimination and the timing of neural responses to pitch change. *Frontiers in Psychology*, 2, 146.
- 936 Goldman, R. I., Wei, C. Y., Philiastides, M. G., Gerson, A. D., Friedman, D., Brown, T. R., et al. (2009). Single-trial discrimination for integrating simultaneous EEG and fMRI: Identifying cortical areas contributing to trial-to-trial variability in the auditory oddball task. *Neuroimage*, 47(1), 136–147.
- 937 Gonsalves, B., & Paller, K. A. (2000). Brain potentials associated with recollective processing of spoken words. *Memory & Cognition*, 28(3), 321–330.
- 938 Haslinger, B., Erhard, P., Altenmüller, E., Hennenlotter, A., Schweiger, M., & Graf von Einsiedel, H. (2004). Reduced recruitment of motor association areas during bimanual coordination in concert pianists. *Human Brain Mapping*, 22(3), 206–215.
- 939 Hasson, U., Nusbaum, H. C., & Small, S. L. (2007). Brain networks subserving the extraction of sentence information and its encoding to memory. *Cerebral Cortex*, 17(12), 2899–2913.
- 940 Holmes, A. P., Blair, R. C., Watson, J. D., & Ford, I. (1996). Nonparametric analysis of statistic images from functional mapping experiments. *Journal of Cerebral Blood Flow and Metabolism*, 16(1), 7–22.
- 941 Jordan, M. I., & Jacobs, R. A. (1994). Hierarchical mixtures of experts and the EM algorithm. *Neural Computation*, 6, 181–214.
- 942 Katahira, K., Abia, D., Masuda, S., & Okanoya, K. (2008). Feedback-based error monitoring processes during musical performance: An ERP study. *Neuroscience Research*, 61(1), 120–128.
- 943 Koelsch, S. (2009). Music-syntactic processing and auditory memory: Similarities and differences between ERAN and MMN. *Psychophysiology*, 46(1), 179–190.
- 944 Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: Brain signatures of semantic processing. *Nature Neuroscience*, 7(3), 302–307.
- 945 Koelsch, S., Schmidt, B. H., & Kansok, J. (2002). Effects of musical expertise on the early right anterior negativity: An event-related brain potential study. *Psychophysiology*, 39(5), 657–663.
- 946 Koelsch, S., Schroger, E., & Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *NeuroReport*, 10(6), 1309–1313.
- 947 Koelsch, S., & Siebel, W. A. (2005). Towards a neural basis of music perception. *Trends in Cognitive Sciences*, 9(12), 578–584.
- 948 Leaver, A. M., Van Lare, J., Zielinski, B., Halpern, A. R., & Rauschecker, J. P. (2009). Brain activation during anticipation of sound sequences. *Journal of Neuroscience*, 29(8), 2477–2485.
- 949 Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1), 1–36.
- 950 Lotto, A. J., Hickok, G. S., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences*, 13(3), 110–114.
- 951 Loui, P., Grent-t-Jong, T., Torpey, D., & Woldorff, M. (2005). Effects of attention on the neural processing of harmonic syntax in Western music. *Brain Research. Cognitive Brain Research*, 25(3), 678–687.
- 952 Loui, P., Li, H. C., & Schlaug, G. (2011). White matter integrity in right hemisphere predicts pitch-related grammar learning. *Neuroimage*, 55(2), 500–507.
- 953 Maidhof, C., Vavatzanidis, N., Prinz, W., Rieger, M., & Koelsch, S. (2010). Processing expectancy violations during music performance and perception: An ERP study. *Journal of Cognitive Neuroscience*, 22(10), 2401–2413.
- 954 Meister, I., Krings, T., Foltys, H., Borojerdi, B., Müller, M., Topper, R., et al. (2005). Effects of long-term practice and task complexity in musicians and nonmusicians performing simple and complex motor tasks: Implications for cortical motor organization. *Human Brain Mapping*, 25(3), 345–352.
- 955 Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, 15(1), 1–25.
- 956 Olmstead, A. J., Viswanathan, N., Aicher, K. A., & Fowler, C. A. (2009). Sentence comprehension affects the dynamics of bimanual coordination: Implications for embodied cognition. *Quarterly Journal of Experimental Psychology (Hove)*, 62(12), 2409–2417.
- 957 Parra, L., Alvino, C., Tang, A., Pearlmuter, B., Yeung, N., Osman, A., et al. (2002). Linear spatial integration for single-trial detection in encephalography. *Neuroimage*, 17(1), 223–230.
- 958 Parra, L. C., Spence, C. D., Gerson, A. D., & Sajda, P. (2005). Recipes for the linear analysis of EEG. *Neuroimage*, 28(2), 326–341.
- 959 Pascual-Marqui, R. D. (2002). Standardized low-resolution brain electromagnetic tomography (sLORETA): Technical details. *Methods & Findings in Experimental & Clinical Pharmacology*, 24(Suppl. D), 5–12.
- 960 Pascual-Marqui, R. D., Esslen, M., Kochi, K., & Lehmann, D. (2002). Functional imaging with low-resolution brain electromagnetic tomography (LORETA): A review. *Methods & Findings in Experimental & Clinical Pharmacology*, 24(Suppl. C), 91–95.
- 961 Pascual-Marqui, R. D., Lehmann, D., Koenig, T., Kochi, K., Merlo, M. C., Hell, D., et al. (1999). Low resolution brain electromagnetic tomography (LORETA) functional imaging in acute, neuroleptic-naïve, first-episode, productive schizophrenia. *Psychiatry Research*, 90(3), 169–179.
- 962 Pfordresher, P. (2006). Coordination of perception and action in music performance. *Advances in Cognitive Psychology*, 2(2–3).
- 963 Pfordresher, P., & Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attention Perception Psychophysics*, 71, 1385–1398.
- 964 Pfordresher, P., & Kulpa, J. D. (2011). The dynamics of disruption from altered auditory feedback: Further evidence for a dissociation of sequencing and

- 1043 timing. *Journal of Experimental Psychology: Human Perception and Performance*,  
1044 37(3), 949–967. 1059
- 1045 Ruiz, M. H., Jabusch, H. C., & Altenmuller, E. (2009). Detecting wrong notes in  
1046 advance: Neuronal correlates of error monitoring in pianists. *Cerebral Cortex*,  
1047 19(11), 2625–2639. 1060
- 1048 Sammler, D., Novembre, G., Koelsch, S., & Keller, P. E. (2013). Syntax in a pianist's  
1049 hand: ERP signatures of "embodied" syntax processing in music. *Cortex*, 49(5),  
1050 1325–1339. 1061
- 1051 Schmithorst, V. J., Holland, S. K., & Plante, E. (2006). Cognitive modules utilized for  
1052 narrative comprehension in children: A functional magnetic resonance imaging  
1053 study. *Neuroimage*, 29(1), 254–266. 1062
- 1054 Sevdalis, V., & Keller, P. E. (2011). Captured by motion: Dance, action understanding,  
1055 and social cognition. *Brain and Cognition*, 77(2), 231–236. 1063
- 1056 Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., et al.  
1057 (2005). Listening to action-related sentences activates fronto-parietal motor  
1058 circuits. *Journal of Cognitive Neuroscience*, 17(2), 273–281. 1064
- Virtue, S., Haberman, J., Clancy, Z., Parrish, T., & Jung Beeman, M. (2006). Neural  
1065 activity of inferences during story comprehension. *Brain Research*, 1084(1),  
1066 104–114. 1067
- Wronka, E., Kaiser, J., & Coenen, A. M. (2012). Neural generators of the auditory  
1068 evoked potential components P3a and P3b. *Acta Neurobiologiae Experimentalis*  
1069 (*Wars*), 72(1), 51–64. 1070
- Zatorre, R. J. (2007). There's more to auditory cortex than meets the ear. *Hearing*  
1071 *Research*, 229(1–2), 24–30. 1072
- Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music:  
1073 Auditory-motor interactions in music perception and production. *Nature*  
1074 *Reviews Neuroscience*, 8(7), 547–558. 1075

UNCORRECTED PROOF