On the specificity of auditory and motor simulation. An fMRI study with expert pianists.

Giacomo Novembre

Supervisor: Shirley-Ann Rueschemeyer

Donders Centre for Cognition, Donders Centre for Cognitive Neuroimaging, Nijmegen, The Netherlands

Correspondence to: Giacomo Novembre; giacomonovembre@hotmail.it

Abstract

The present study investigates the neural correlates of visual-driven auditory and motor representations in expert musicians. Expert pianists were presented with videos showing one hand playing chord sequences on a mute keyboard. The last chord of the sequence was either congruent or incongruent to the musical context with regard to (1) auditory expectation (the harmony of the anticipated chord progression) or (2) visual-motor expectation (the set of fingers normally used to play harmonic chords). Both the incongruent conditions were not wrong chords per se, but they were incompatible with the context on the basis of their position along the chord progression. In a behavioural study, we showed that expert pianists were able to significantly discriminate these incongruent conditions, while non-musicians performed at chance-level. This suggests that, although the videos were silent, experts were able to extract very specific auditory and motor information from the visual presentation of the melodies. In order to understand how expert musicians performed this task in the absence of auditory input, we measured the neurophysiological correlates of this ability using fMRI. We hypothesized that auditory and motor simulation of musical actions drove performance in this task. As we expected, acoustically violated trials activated secondary auditory cortex (Middle Temporal Gyrus), suggesting that auditory simulation underlies the establishment of auditory expectancy. Likewise, action-violated trials activated the cerebral motor system (Middle Frontal Gyrus and Inferior Parietal Lobe), suggesting that motor simulation underlies the establishment of visual-motor expectancy. In conclusion, this study provides evidence in favour of specific auditory and motor simulation mechanisms. Additionally, it provides new insights into how musical expectations are derived in expert musicians by showing that expert musicians rely on simulation while observing musical performance. These simulation mechanisms have been speculated to serve action prediction, a crucial component of joint musical actions.

Keywords: action observation, auditory simulation, motor simulation, musical expertise, fMRI.

Introduction

Music cannot be produced without movement. Intuitively, music could be defined as the acoustic account of very complex sequences of movements. However, from musicians' perspective, this frame is reductive, because each sound also has a visual component, i.e., how a specific action, that produces a specific sound, looks like. This visual component could work as a form of non-verbal communication between players in several situations. For example, it could facilitate them to learn to play an instrument from a teacher; to coordinate with another during joint performance; or to synchronize action in music improvisation (Vines, 2006). The present study investigates the specificity of visual-driven acoustic and motor representations in expert musicians.

A number of studies recently reported a high degree of coupling between action and auditory perception in expert musicians (for a review see Zatorre et al., 2007). These findings imply that musicians process musical information by incorporating more modalities than people who can still enjoy music but cannot play it. For instance, in an fMRI study, Bangert et al. (2006) compared the brain responses of expert pianists to those of non-musicians while playing a mute piano keyboard in the scanner. This motor-task activated secondary auditory cortex, including superior and middle temporal gyrus (STG and MTG), in pianists but not in non-musicians. This finding implies that pianists had an auditory representation of the musical action, even if they did not listen to any sound. Along the same line, in an fMRI study, Haslinger et al. (2005) examined if expert pianists had different brain activation from non-musicians during passive watching of one hand playing on a mute keyboard. This action observation task also activated secondary auditory cortex in pianists but not in non-musicians, implying that visual perception can also lead to an auditory representation of a musical action.

It is interesting to note that the pattern of activations found by Bangert et al. (2006) in expert musicians for a motor task was strictly similar to those found by Haslinger et al. (2005) during observation of similar actions. In particular, premotor cortex (PMC) and inferior parietal cortex (IPC) were active in both tasks, indicating that visual observation of mute piano playing recruited the same neural mechanisms involved in the execution of a similar action. Additionally, secondary auditory cortex was active in

both the experiments, suggesting that this visual-motor matching system might be an audio-motor system at the same time. A similar conclusion was drawn by Gazzola et al. (2006), who compared a network of areas active during both execution and auditory presentation of sound-related actions (e.g. ripping paper, breaking peanuts) with the visual presentation of the same actions. Interestingly, Gazzola et al. (2006) reported a group of areas active across modalities including Inferior Frontal Gyrus (IFG) and MTG (see Gazzola et al., 2006, supplementary data). This implies that these motor and auditory representations of visually perceived sound-related actions are common to everybody and especially developed in skilled musicians as a result of long-term training.

In order to interpret the nature of these visual-driven auditory and motor activations, one could claim that they consist of motor and auditory simulations respectively. Motor simulation consists of recruiting one's own motor system to represent and understand the actions of others (Gallese & Goldman, 1998; Gallese, 2007), while auditory simulation consists of recruiting one's own auditory system to represent and understand sound-related actions. However, a number of authors have recently challenged this interpretation (Mahon & Caramazza, 2008; Hickok, 2009) and pointed out that these activations might consist of general associations. Specifically, Mahon & Caramazza (2008) have proposed that these activations could be mediated by "abstract" or "symbolic" representations of the concepts representing actions. This means that, e.g., observing one hand playing a keyboard would first induce a conceptual processing that would then serve to ground abstract and symbolic representations in the rich sensory and motor content. In other words, there are two ways in which such activations may come about. Firstly, they could be semantically mediated (general associations). Secondly, they could consist of a direct transformation of perception into action (simulations).

The present study aimed at distinguishing between these two interpretations by looking at the *specificity* of these activations. We supposed this specificity to be a critical feature for determining whether these activations consist of specific simulations or conceptual associations. In fact, general associations cannot reflect specific sounds. Therefore, we assumed that showing that this activity reflects specific sounds or specific actions, rather than sound or action in general, would be good evidence in favor of specific simulation mechanisms.

In order to address this experimental question, we created a set of musical stimuli composed of five-chord sequences. More precisely, the first four chords established a musical context toward the end of the sequence while the last chord could be congruent or incongruent (in terms of classical harmony, see experimental methods) with the previous musical context. Previous ERP and fMRI studies showed that acoustic presentation of these incongruent chords elicits a specific pattern of activations (Koelsh & Friederici, 2003; Koelsh et al., 2005). In particular, Koelsh et al. (2005) showed that incongruent chords presented at the end of a musical sequence elicit a stronger BOLD signal in the secondary auditory cortex (STG and MTG) and in IFG. Hence, if visually perceived musical actions are processed by way of auditory simulations, we expected 1) that expert pianists were able to detect these incongruencies by only watching the mute musical actions and 2) that the neural mechanisms active during visual detection of errors were comparable to those seen in previous studies using auditory stimuli.

Subjects were shown with videos presenting one hand playing on a mute keyboard. In a control condition (Congruent condition, Figure 1A) the last chord was congruent with the previous context. In a first experimental condition aimed to test the specificity of auditory simulation, the last chord was not congruent in terms of classical harmony (Audio incongruent condition, Figure 1B). In a second experimental condition aimed to test the specificity of motor simulation, the last chord was congruent in terms of classical harmony but not congruent in terms of the fingering used to play a chord (Motor incongruent condition, Figure 1C). The incongruent chords used in the experimental conditions were not wrong chords per se, but they were incompatible with the context on the basis of their position along the chord progression. This means that piano players are familiar with these chords, which are part of their motor repertoire, but they are not used to play them at the end of a melody. Therefore, we expected these incongruencies to elicit a surprising effect due to the mismatch between piano players' auditory and motor expectations and visual observation.

In a behavioral study, we tested whether expert pianists were able to successfully perform a detection task. Moreover, in order to be sure that a successful performance of the task relied on a specific musical expertise, we included a control group of non-musicians. Subjects were visually shown mute sequences and asked to detect auditory and motor incongruencies along the chord progression. Results

indicated that musicians were able to use action observation to elicit specific auditory and motor information. However, it was not possible from this behavioral study to clarify whether this ability relied on a semantic strategy or on specific auditory and motor simulation. For this reason, we ran a follow-up fMRI investigation aimed to explore the neural correlates supporting this ability.

Functional MRI measurements were expected to highlight the neural correlates of auditory and motor simulation. In particular, if auditory simulation underlies the establishment of auditory expectancy in musicians, then we expected to see activation in secondary auditory cortex (STG and MTG) during perception of acoustically violated trials. Likewise, if motor simulation underlies the establishment of visual-motor expectancy, then we expected to see activation in the cerebral motor system (i.e., inferior parietal cortex, premotor cortex) for action-violated trials.

Behavioral Study

Experimental Methods

Participants

We investigated 10 professional pianists (5 women and 5 men, mean age 24.2 +- 3.08) and 10 musically-naïve subjects (8 women and 2 men, mean age 20.5 +- 2.8). Musical professionalism was defined as 1) having continuously played piano for the last 10 years and 2) playing with an average of 20 hours per week. Musicians' mean history of piano practice was 15.2 +- 3.3 years. Musically naïve controls had to have never played a musical instrument before and have never received any musical teaching beyond normal school education. All subjects were naïve with regards to the goal of the study. No subject had any history of neurological disease. All subjects gave written consent according to the guidelines of the local ethics committee (CMO Committee on Research Involving Humans subjects, region Arnhem-Nijmegen, the Netherlands).

Stimuli

Stimuli were videos presenting a right hand playing five-chord sequences on a mute keyboard. Presentation time was 2000 ms for each chord plus 1000 ms at the beginning of each video resulting in a total duration of 11000 ms for each sequence. Chord sequences were played on an EZ200 Yamaha Keyboard provided with an LED system that was supposed to help subjects understanding which specific keys were being pressed. In particular, each button was provided with a red light that lit up when pressed. Lights' intensity was homogenous within black or white buttons, but different between black and white buttons.

A total of 120 unique chord sequences were created for this experiment: 40 were regular (Congruent), 40 were irregular regarding harmonic aspects (Audio Incongruent) and 40 were irregular regarding motoric aspects (Motor Incongruent) (see Figure 1, right part). All chords included in the sequences consisted of three tones. The first chord was always the tonic of the following sequence. Chords at the second position were tonic, mediant, or subdominant. Chords at the third position were subdominant, dominant, or dominant six–fourth chords. Chords at the fourth position were dominant seventh chords presented in root, first and third inversion. The chord at the fifth position differed between three conditions: a tonic chord (Congruent), a Neapolitan chord (Audio Incongruent) or a tonic chord played with alternative fingering (Motor Incongruent). Tonic chords were presented in root position, first and second inversions (see Figure 1, left part).

a) Audio incongruent chords

A Neapolitan chord is a minor subdominant with a diminished sixth instead of a fifth. Although classical Neapolitan chords most often occur in first inversion, they were presented in both first and second inversion according to the harmony principle of chord progression. Four major keys with maximum two key signatures were included (C, D, F, G). Chord sequences were presented with different melodic outlines (e.g., starting with the third, the fifth, or the octave in the top voice).

b) Motor incongruent chords

There are commonly used fingerings for playing chords in the basic classical piano playing, e.g., root position of tonic is played by thumb, middle finger, and little finger (right hand). However, the same chord could also be played using thumb, index and ring fingers. Such alternative fingerings are typically chosen when preparing to play the next chord. Therefore, ending the chord progression with alternative fingerings is less common.

Task and Procedure

Participants were presented with a total of 144 experimental stimuli. Stimuli belonged to one of the three conditions: (1) sequences ending with a congruent chord (Congruent), (2) sequences ending with a not harmonic chord (Audio incongruent) and (3) sequences ending with an unusual fingering (Motor incongruent). In order to counterbalance the probability for the participants to watch a congruent or an incongruent video, each congruent trial was presented two times, for a total of 72 congruent sequences and 72 incongruent sequences (36 Audio and 36 Motor respectively). Congruent videos presented for the second time were filtered out from the analysis to exclude any involvement of memory. The order of stimuli presentation was randomized individually for each participant. All participants saw all experimental stimuli.

A trial consisted of visual presentation of a sequence. At the beginning of each trial, a fixation cross appeared on the screen for 500 ms. Directly following the fixation cross the video started.

After the presentation of each video, subjects were asked to judge the congruency of the chord progression by pressing a button after the last chord of each sequence: one button for the congruent sequence ending (Congruent), and another button for the incongruent ending (Audio or Motor incongruent). Subjects were not explicitly told that incongruent chords were always placed at the last position.

Before the experiment started, subjects performed a training session that consisted of the presentation of eight videos: four congruent sequences and four incongruent sequences (two Audio and two Motor sequences). During the training session, but not in the experimental session, subjects received feedback on their

response and were allowed to watch the same video twice. The videos presented during the training were not presented again during the experiment.

Post-Experiment questionnaire

Professional pianists and naïve subjects were asked to fill in a questionnaire after the experiment. The aim of this questionnaire was to 1) define which parameters of players' expertise were better predictors of their performance, 2) to control the group of non-musicians for any indirect experience with music, and 3) to understand which strategy was adopted by subjects to perform the task.

Professional pianists were asked to specify how many years of experience they had (Experience), at which age they started to play piano (Onset Age), how many hours on average they used to play piano (per week) (Frequency) and (if applicable) how many hours on average they used to teach piano (Teaching).

Naïve subjects where asked to describe any indirect experience with piano playing or, in general, with music. Both groups were asked to describe the strategy they used to perform the task. Specifically, expert players were asked if they relied on their theoretical knowledge of the music rules or on imagining playing the sequences themselves.

Results

Behavioral performance

Mean performance rates of a group of 10 pianists and a control group of 10 non-musicians were compared to the chance level using one-sample t tests.

Expert pianists performed significantly above chance level (50%) in the audio condition (Performance rates: mean = 97.90, SD = 3.01, t-value = 50, DF = 9, $p<.001$) and in the motor condition (Performance rates: mean = 89.60, SD = 13.51, t-value = 50, DF = 9, $p<.001$) (Figure 2, Panels A and B, left columns).

Mean performance rates of non-musicians did not significantly differ from chance level in both the audio condition (M = 54.30%, SD = 20.35, t-value = 50, DF = 9,

p>.1) and the motor condition (M = 47.20%, SD = 17.69, t-value = 50, DF = 9, *p*>.1) (Figure 2, A and B, right columns).

Post-Experiment questionnaire report

Questionnaire's data regarding expert musicians is shown in Table 1.

Linear regression analysis between performance rate in the audio condition and in the motor condition (dependent variables: Audio, Motor) and four possible predictors of their performance rate (independent variables: Experience, Onset Age, Frequency, Teaching) were run. Results show a significant interaction between the number of hours players used to teach piano and their performance rate in the audio condition (R square = 0.589, *p*<.05). The other interactions were not significant (*p*>.1).

Eight out of 10 pianists reported that imagining playing the sequences themselves facilitated their performance during the task. Other two subjects reported to also have used their theoretical knowledge of musical rules.

None of the non-musicians reported to have ever had any noteworthy experience with music beyond normal school education and, in general, with music performance. None of these subjects figured out a successful strategy to perform the task.

Discussion

Behavioral performance

The present behavioral study tested whether expert pianists were able to successfully detect very specific auditory and motor incongruencies by only watching one hand playing a mute piano keyboard. Moreover, in order to ensure that a successful performance of this task relied on specific musical expertise, a control group of non-musicians was also tested. The high performance rates of pianists in both conditions suggested that they were able to perform the task successfully. Furthermore, the chance-level performance of non-musicians indicated that musical expertise is required to carry out the task. Results indicated that musicians were able to use action observation to elicit specific auditory and motor information. However, it was not possible to clarify if the ability to detect auditory violated trials relied on a

specific simulation process or on a semantic strategy. In fact, participants might have used two different strategies to perform the task. On one hand, they could have simulated the specific sounds generated by the visually perceived actions. On the other, they could have employed a semantic strategy (i.e. naming chord by chord) and detected harmonic violations by making use of their theoretical understanding of musical rules. Therefore, in order to distinguish between these two possible interpretations, we ran a follow-up fMRI investigation aimed to explore the neural correlates supporting this ability (see fMRI study below).

Performance-Expertise interactions

Results indicate that players who used to teach more often were better in detecting auditory incongruencies. This further developed ability might be due to their regular practice in observing others' playing and checking the correctness of others' performance.

The other variables for measuring pianists' expertise (Experience, Onset Age, Frequency) were no reliable predictors of their performance. The cause of this unsuccessful analysis might be related to our test material, which was relatively easy for the specific population tested in this experiment. This is demonstrated by the fact that most subjects' had ceiling performance. Therefore, in order to further investigate which one of these factors can explain variance in the task performance, a follow up study could either utilize different stimuli or test the same stimuli we used with a population of less skilled pianists.

Finally, the fact that most pianists found imagining playing the sequences an efficient strategy to perform the task offers a positive indication of a simulation process.

Functional MRI Study

Experimental Methods

Participants

For this experiment, the same group of 10 expert pianists who took part in the behavioral experiment was recruited.

Stimuli

For this experiment, the same stimuli as in the behavioral experiment were used.

Task and Procedure

In this fMRI experiment, task and procedure were slightly different to those of the behavioral experiment. In particular, in addition to the 144 critical trials used in the behavioral experiment (i.e. 72 Congruent sequences and 72 incongruent sequences: 36 Audio sequence and 36 Motor sequences), 18 null events were presented in order to include an additional control condition.

A trial consisted of visual presentation of a sequence (or in the case of null events, it was a presentation of black screen). At the beginning of each trial, a variable jitter time of 3000, 4000 or 5000 ms was included in order to improve the sampling rate of BOLD signal. Following the jitter, a white fixation cross appeared on the screen for 500 ms. Directly following the fixation cross the video started.

After the presentation of each video, subjects were asked to judge the congruency of the chord progression by pressing a button after the last chord of each sequence: one button for the congruent sequence ending (Congruent) using the right hand's index finger, and another button for the incongruent ending (Audio or Motor incongruent) using the middle finger of the same hand.

Before the experiment started, subjects performed the same training session performed as in the behavioral study.

fMRI data acquisition

Functional images were acquired on a Siemens TRIO 3.0 T MRI system (Siemens, Erlangen, Germany) equipped with echo planar imaging (EPI) capabilities, using a birdcage head coil for radio frequency transmission and signal reception. Blood oxygenation level-dependent (BOLD) sensitive functional images were

acquired using a single shot gradient EPI-sequence (TE/TR = 30/2000 ms; 31 axial slices in ascending order, voxel size = 3.5 x 3.5 x 3.5). High-resolution anatomical images were acquired using a MPRAGE sequence (TE = 3.03; voxel size = 1 mm x 1 mm x 1 mm, 192 sagittal slices, FOV = 256).

fMRI data analysis

Functional data were pre-processed and analyzed with SPM5 (Statistical Parametric Mapping, www.fil.ion.ucl.ac.uk/spm). Rigid body registration along 3 translation and 3 rotations was applied to correct for small head movements. Subsequently the time series for each voxel was realigned temporally to acquisition of the middle slice (slice 16), to correct for slice timing acquisition delays. Images were normalized to a standard EPI template centered in MNI space and re-sampled at a isotropic voxel size of 2 mm. Low-frequency signal changes and base-line drifts were removed by applying a temporal highpass filter to remove frequencies lower than 1/128 Hz. The normalized images were smoothed with an isotropic 10 mm full-width-at-half-maximum (FWHM) Gaussian kernel. The ensuing pre-processed fMRI time series were analyzed on a subject-by-subject basis using an event-related approach in the context of the General Linear Model (GLM) with regressors for each condition (observation of a congruent chord, observation of an audio incongruent chord, observation of a motor incongruent chord, observation of null-event) convolved with a canonical hemodynamic response function. The parameters from the motion correction algorithm were included in the model as effects of no interest. Two different contrasts were defined comparing observation of audio incongruent chord vs. observation of congruent chord (Audio – Congr) and observation of motor incongruent chord vs. observation of congruent chord (Motor – Congr). This resulted in a total of 2 contrast images per subject.

Because individual functional datasets had been aligned to the standard stereotactic reference space, a group analysis based on the contrast images could be performed. Single-participant contrast images were entered into a second-level random effects analysis for the critical contrast of interest. The group analysis consisted of a one-sample t-test across the contrast images of all subjects that indicated whether observed differences between conditions were significantly distinct

from zero. To protect against false positive activation, a double threshold was applied, by which only voxels with a $p < .005$, uncorrected and a volume exceeding 65 voxels were considered (Slotnick et al., 2003).

Results

Differential activation during the detection of auditory incongruent chords compared to the detection of congruent chords.

A list of significant activations can be seen in Table 2 and Figure 3 (Panel A).

We determined activation differences when comparing observation of one hand playing a regular chord (Congruent condition) to the observation of one hand playing a not harmonic chord (Audio incongruent). For observation of not harmonic chords, participants showed significantly stronger activations in the following areas.

Two big blobs of activation were seen in the occipital cortex consisting of right and left middle and inferior occipital gyrus. The first extended from right lingual gyrus and right calcarinus sulcus to right occipital-temporal scissure, including right cerebellum (crus 1). The second extended from left descendents gyrus to left anterior occipital sulcus, including left cerebellum (crus 1).

In the right parietal cortex, two blobs of activity were seen: the first one in the superior parietal lobe, extending from right angular gyrus to the anterior and superior occipital sulcus, the second in the inferior parietal cortex (PGp). In the left parietal cortex, one blob of activation including superior and inferior parietal lobule was seen. This blob extended from the secondary intermediate sulcus to the intraparietal sulcus, and laterally to the angular gyrus.

Over the temporal cortex, activation was seen bilaterally in the middle temporal gyrus. In the left hemisphere, this activation extended laterally to the proximity of the superior temporal sulcus.

Middle frontal gyrus was activated bilaterally. Over the right hemisphere, precentral gyrus was also active. Over the left hemisphere, left inferior frontal gyrus (pars opercularis) was activated.

Differential activation during the detection of motor incongruent chords compared to the detection of congruent chords.

A list of significant activations can be seen in Table 3 and Figure 3 (Panel B).

We determined activation differences when comparing observation of one hand playing a regular chord (Congruent condition) to the observation of one hand playing a chord with irregular fingering (Motor incongruent). For observation of chords with irregular fingering, participants showed significantly stronger activations in the following areas.

In the left frontal lobe, three blobs of activity were seen. The first one was localized in the superior frontal gyrus. The second one was localized in the inferior frontal gyrus, corresponding to pars angularis and pars triangularis. The third one was located in the middle frontal gyrus.

In the left inferior parietal cortex, a blob of activation corresponding to the left angular gyrus (PGa) was also seen.

Over the right hemisphere, a blob extending from the right occipital cortex until the right occipital-temporal fissure was seen.

Discussion

The contrast between regular chords (Congruent condition) and not-harmonic chords (Audio condition) indicated that the latter chords elicited greater activation than regular chords within the secondary auditory cortex. In particular, activation was seen in the Middle Temporal Gyrus (MTG). This activation indicates that action observation was transformed in a very specific auditory simulation by means of a cross-modal processing. Indeed, a number of studies that used the same auditory violations as in the present study (Koelsh & Friederici, 2003; Koelsh et al., 2005) also reported stronger secondary auditory activations (in STG and MTG) during acoustic detection of not-harmonic chords. This implies that visual observation of mute piano playing can result in auditory representations that reflect specific notes rather than general sound.

The contrast between regular chords (Congruent condition) and chords played with a not-congruent fingering (Motor condition) indicated that the latter chords

elicited greater activation than regular chords within the motor system. In particular, activation was seen in the middle frontal gyrus (lateral premotor cortex, PMC), and inferior parietal cortex (IPC). Results are consistent with previous literature on motor simulation (Manthey et al., 2002; Koelewijn et al., 2008) and indicate the specific way in which these actions are simulated.

General Discussion

Specificity of auditory simulation

Our results are consistent with previous studies showing that executing sound-related actions in the absence of sound, as well as listening to the sound of the same actions, activates a common pattern of areas including MTG (Bangert et al., 2006; Gazzola et al., 2006). Moreover, it is important to highlight that this area shows selective activation for action-related sounds compared to general environmental sounds (Gazzola et al., 2006). For this reason, we hypothesize that MTG activation reflects a *motor-based* auditory simulation. This interpretation is also consistent with 1) fMRI results showing a number of motor areas (including PMC and IPC) active during observation of Audio incongruent chords and 2) results of the post-experiment questionnaire (where eight subjects of ten revealed that imagining playing the observed sequences themselves facilitated their performance in the task). Piano players are more trained to perform music than to visually perceive it, consequently it is reasonable to assume that their motor representation of music performance facilitated detection of acoustic incongruencies better than their visual representation.

In a recent fMRI study using an unconscious semantic priming paradigm, Gelati et al. (2007) provided divergent evidence on the nature of MTG activations. In this study, auditory presentations of action-related sounds were preceded by word-primes that could be congruent (i.e. having a semantic relationship) or not with the following sound. The modulation of left IFG and left MTG by the congruency of the prime-target semantic relationship (i.e. activation was higher for action-related sounds preceded by congruent than incongruent primes) suggested that these regions code only the specific feature that is eventually shared by the prime word and the sound, i.e., the semantic meaning of that action. Although these results clearly show that

semantic associations can drive a modulation in IFG and MTG, this evidence does not necessarily imply that our activations also reflect a semantic association (i.e. naming the chords perceived) because of two following reasons. Firstly, because post-experiment questionnaire suggested that most of the subjects employed a simulation-strategy in order to perform the task. Secondly, because fMRI results show that auditory violated trials activated both left and right MTG. In Gelati's et al. (2006) study, right MTG did not showed the main effect of prime and this result was interpreted as evidence against the involvement of right MTG in coding actions at semantic level. Therefore, the fact that right MTG showed additional activation in response to our auditory incongruencies, and not during semantic associations, supports our hypothesis of a specific auditory simulation.

It is also necessary to compare our findings with those in other two similar studies that previously investigated auditory representations in piano players during observation of mute piano performance (Hasegawa et al., 2004; Haslinger et al., 2005). These studies also reported secondary auditory activity elicited in response to watching one hand playing a mute keyboard, but this activation was seen in STG rather than MTG. We believe the reason of this inconsistency to rely on the fact that our experiment investigated pianists' ability to detect very specific auditory incongruencies, a process that did not occur in the other two studies. In fact, our task required subjects to detect not-harmonic chords embedded in musical sequences, while the other studies investigated pianists' passive watching (Haslinger et al., 2005) or their ability to recognize known melodies (Hasegawa et al., 2004).

In particular, Haslinger et al. (2005) compared meaningful musical actions (silent hand playing the piano) with its meaningless motor equivalent (sequential finger to thumb opposition movement) and showed that auditory representations, consisting of STG activity, were selective for sound-related actions. However, no behavioral evidence showed that participants were engaged in a specific auditory representation of the sound-related actions (i.e. representing specific notes) and thus this activation could only reflect general acoustic association.

Hasegawa et al. (2004) compared expert pianists, less trained pianists and naïve subjects while they were shown mute piano execution of familiar music, not familiar music and random sequences (sequence that do not contain any musical structure by listening to them). Subjects were asked whether they were able to recognize the played melodies or not. Behavioral results indicated that unfamiliar pieces of music

(the only comparable condition to our experimental material) were not well identified by any of the trained groups (3.3 +- 0.3% for less trained, 18.5 +- 0.9% for well trained, difference not statistically significant, naïve: 0%) indicating that trained subjects were not able to extract very specific auditory information from unknown melodies. However, fMRI measurements showed that familiar music, unfamiliar music and random movement equally activated left STG in expert pianists, whereas the same region was less activated in all conditions in the naïve and less trained groups. According to Hasegawa et al. (2004), this activation reflects the mapping of complex sequence structures of hand movements onto the sequences of sounds, but no information about the specificity of these auditory representations is reported. Because expert musicians hardly detected unfamiliar music pieces, it would be pretentious to claim that they were able to perform such a difficult task as mapping the specific sounds of random sequences (in particular, because random sequences did not contain any musical structure by listening to them). Therefore, equal STG activation across the three conditions suggests that this area reflected general auditory association, due to long-term training, rather than a specific auditory simulation. Oppositely, if STG activation had been consistent with the behavioral performance (e.g. showing different activation during such a unfeasible task as mapping random sequences), then it could be argued that this activation reflects a specific simulation.

In conclusion, although previous studies localized the neural correlates of auditory associations in a different region of the secondary auditory cortex (i.e. STG instead of MTG) (Hasegawa et al., 2004; Haslinger et al., 2005), this inconsistency can be explained by the different task and by the different stimuli used in the present investigation. Moreover, although activations do not overlap, they are proximal to one another and they are both in areas relevant to auditory processing,

Specificity of motor simulation

Our results replicated pre-existing evidence in favor of specific execution-observation interactions in monkeys' and human beings' brain, and provided additional evidence in favor of a specific motor simulation mechanism.

Neurons in the PMC of monkey respond during the execution of the monkey's own actions and while the monkey observes (di Pellegrino et al., 1992; Keysers et al., 2003) or hears (Kohler et al., 2002; Keysers et al., 2003) other individuals perform similar actions. Neurons with similar properties have also been found in the rostral inferior parietal lobule of monkey (Fogassi et al., 2005; Fujii et al., 2007). Several neuroimaging studies have shown that similar activations, including PMC and IPC, also occur in human beings' brain during execution and observation of similar actions (Iacoboni et al., 1999; Molnar-Szakacs et al., 2006). Moreover, other studies have showed that specific motor expertise can modulate activation in the motor cortex during observation of actions that are well established in one's motor repertoire (Haslinger et al., 2005; Calvo-Merino et al., 2005 and 2006; Van Elk et al., 2008).

These co-activations have been interpreted to reflect the neural substrates that would allow individuals to understand others' actions by associating them with us through a simulation (Gallese & Goldman, 1998; Gallese, 2007). This hypothesis has been tested by a number of studies that compared brain responses during the observation of meaningful and meaningless actions (Manthey et al., 2002; Koelewijn et al., 2008). Results showed that activations within the motor system are modulated by the correctness of the observed actions, suggesting that the motor system probably contribute to provide a "natural basis for understanding and interpreting the actions conveyed by others" (Koelewijn, 2008, p. 774).

The present study reproduced existing evidence supporting the view that this network of areas is active when interpreting the meaning of the observed action. In fact, motor incongruencies were selective violations of the way in which goal-directed actions were executed. The additional activation of the motor system for this condition reflects participants' effort to understand the meaning (expressed in an unusual way) of the observed action. We exclude this additional activation to be semantically mediated (Mahon & Caramazza, 2008) because the chords played in the Motor condition and in the Congruent condition were the same (i.e. the only different feature was the set of fingers used). Therefore, if task-performance had been driven by a semantic strategy (i.e. naming of the chords), this activation would have been expected to be equal between the conditions. However, this did not happen, implying that the additional activation reflects a specific motor simulation (i.e. simulation of a different set of fingers).

A potential confound in the audio condition

It is important to underline that, from a visual perspective, all the audio incongruent chords were recognizable by means of more of black keys being played. For this reason, it would have been sensible to counterbalance the amount of black keys involved in every sequence. However, during the preparation of the stimuli we realized that such a set of stimuli would have made the task very difficult. Thus, for the feasibility of the task, stimuli were not counterbalanced. It could be argued that the larger number of black keys could have biased the results of this experiment by facilitating musicians in their task (i.e. if many black keys are involved, then the sequence is not congruent). However, there are two good reasons why this was not the case. Firstly, because the group of non-musicians tested in the behavioral study was not able to detect the incongruencies, implying that the difference was not very prominent. Secondly, because none of the experts reported (in the post experiment questionnaire) the difference in number of black and white keys to be a cue to perform the task. Most participants did not even note there was a significant difference.

A potential confound in the motor condition

A number of studies recently showed that music perception might induce involuntary movements (Haueisen & Knosche, 2001). For this reason, it could be argued that involuntary movement induced by action observation caused the motor activation discussed above. To make sure that even minimal muscle activity did not occur, a simultaneous EMG of the finger muscles of the right hand would have been necessary. However, this has not been done in the current fMRI study for the following two reasons. Firstly, because in a previous study, using similar stimuli and comparable subjects group, a co-registered EMG showed no deflection during the presentation of a right hand playing on the piano (Haslinger et al., 2005). Secondly, because in our analysis, we only ran specific contrasts between conditions that all included observation of motor actions. Thus, if involuntary movements were responsible for the activation of the cerebral motor system, then it should be reported that incongruent chords elicited a stronger response of the motor system engaged in involuntary action execution. From our perspective, such result would still support a

direct link between the specific actions observed and the respective motor simulations because an incongruent chord would still lead to a stronger activation of the motor system. Consequently, this would support our hypothesis on a specific motor simulation.

Implications

Although experimental conditions were specific violations regarding auditory and motor aspects respectively, it is important to note that they elicit a common activation pattern of motor areas consisting of: left inferior parietal cortex and left premotor cortex. This suggests that auditory violations are also motor violations at the same time and indicates that the motor system played a crucial role during detection of both incongruencies.

Activation within the motor system in response to observed actions has been attributed to action prediction (Umiltà et al., 2001; Kilner et al., 2004; Fogassi et al., 2005). More specifically, Kilner et al. (2007) claimed that this system is active when generating predicted kinematics from our own motor commands (during motor execution) as well as when observing kinematics produced by others (during perceptual inference). This means that given an expectation of the goal of the person we are observing, we can predict their motor commands. Given the motor commands, we can predict the kinematics on the basis of our own action system. The comparison of the predicted kinematics with the observed kinematics generates an error. This prediction error is used to update our representation of the person's motor commands. Similarly, the inferred goals are updated by a process of minimizing the error between the predicted and the inferred motor command (Kilner, 2007).

Following this hypothesis, a more detailed physiological description of musicians' performance during our experiment can be drawn. In particular, it could be argued that presentation of mute piano sequences automatically activated an internally generated kinematics representation aimed to predict the goal of motor commands of the observed person. In other words, subjects' simulation of the observed motor performance allowed them to make up specific expectations regarding sensory consequences: visual-motor (the way you expect a final chord to be played like) and

audio-motor (the sound you expect a final chord to sound like). Whenever the comparison between the observed actions and subjects' expectations did not match, the elicited error was larger and resulted in a stronger BOLD response within the motor or the sensory system respectively (Friston, 2005).

The ability to predict other people intentions during action observation has been discussed to play a crucial role in musical interactions (Keller et al., 2007 and 2008) and in general joint actions (Sebanz et al., 2006). Indeed, to interact successfully with others, it is often not sufficient to only understand what they are doing at a given moment in time. Instead, being able to predict outcomes of others' actions and knowing what others are going to do next is crucial (Verfaillie et al., 2002; Sebanz et al., 2006). These situations are often found in music production, where understanding other players' actions, predicting their intentions, and co-acting together must all be crucial components. For example, musicians must synchronize their actions with others when playing in a group (e.g. orchestra, ensemble). Last but not least, expert players are even able to improvise music within a group of players, an action that also requires inferential processing. For this reason, we suggest that specific auditory and motor simulations are the mechanisms contributing to specific auditory and motor predictions while observing musical performance. These predictions might be the basis of successful musical interaction. Future studies should further investigate this topic more specifically.

Conclusion

In two experiments, we investigated the neurophysiological mechanisms that allow expert musicians to understand the meaning of visually perceived musical actions. This was done by looking at the specificity of auditory and motor representations elicited in response to visual observation of mute piano performance. Behavioral results showed that expert musicians were able to extract very specific auditory and motor information by only observing these mute actions, while non-expert were not, implying that musical expertise drives performance in this task. However, it was not possible from a behavioral study to establish whether successful musicians' performance was driven by semantic strategy (i.e. a theoretical understanding of musical rules) or by specific auditory and motor simulation. Functional MRI results

provide evidence in favor of specific auditory and motor simulations in expert players, as demonstrated by the fact that regions specialized either for actions or for environmental events show a selective activation for specific incongruencies. These specific simulation mechanisms have been speculated to serve action prediction: an important component of general joint actions, specifically of music.

Acknowledgments

I would like to thank Shirley-Ann Rueschemeyer for her thorough supervision, Makiko Sadakata for her advices in music and for creating and recording all the experimental stimuli, and Harold Bekkering's group for the support provided. I would also like to thank Sasha Ondobaka for his technical assistance, Gerard van Oijen for creating the LED system of the piano keyboard used in this study and Kanako Inque for advertising this experiment in several Dutch conservatories.

References

Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., Heinze, H.J., Altenmuller, E., (2006). Shared networks for auditory and motor processing in professional pianists: evidence from fMRI conjunction. *NeuroImage* 30, 917–926.

Baumann S., Koeneke S., Schmidt C. F., Meyer M., Lutz., Jancke., (2007). A network for audio-motor coordination in skilled pianists and non-musicians. *Brain Research*, 61, 65-78.

Buccino, G., Binkofski, F., Fink, G.R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R.J., Zilles, K., Rizzolatti, G., Freund, H.J., (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *Eur. J. Neurosci*, 13, 400-404.

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. Cereb. *Cortex 15,* (8), 1243–1249.

Calvo-Merino B., Grèzes J., Glaser D. E., Passingham R.E., Haggard P. (2006). Seeing or Doing? Influence of Visual and Motor Familiarity in Action Observation. *Current Biology*, Volume 16, Issue 22, 2277.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., Rizzolatti, G., (1992). Understanding motor events: a neurophysiological study. *Exp. Brain Res.*, 91, 176-

180.

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal lobe: From action organization to intention understanding. *Science*, 308, 662–667.

Friston K., (2005). A theory of cortical responses. *Phil. Trans. R. Soc. 360,* 815-836.

Fujii N, Hihara S, Iriki A. (2007). Dynamic Social Adaptation of Motion-Related Neurons in Primate Parietal Cortex. *Plos* 2, e397.

Gallese, V., and Goldman, A. (1998). Mirror neurons and the simulation theory of mind reading. *Trends Cogn. Sci*. 2, 493-501.

Gallese V., (2007). Before and below 'theory of mind': embodied simulation and the neural correlates of social cognition. *Philosophical transactions*, vol. 362, no. 1480, 659-669.

Gazzola, V., Aziz-Zadeh, L., Keysers, C., (2006). Empathy and the somatotopic auditory mirror system in humans. Curr. Biol., 16, 1824-1829.

Gelati, G., Committeri, G., Spitoni, G., Aprile, T., Di Russo, F., Pitzalis, S., Pizzamiglio, L., (2008). A Selective Representation of the Meaning of Actions in the Auditory Mirror System. *NeuroImage* 40, 1274–1286.

Hasegawa, T., Matsuki, K., Ueno, T., Maeda, Y., Matsue, Y., Konishi, Y., Sadato, N., (2004). Learned audio-visual cross-modal associations in observed piano playing activate the left planum temporale. An fMRI study. *Cogn. Brain Res.,* 20, 510 – 518.

Haslinger, B., Erhard, P., Altenmuller, E., Schroeder, U., Boecker, H., and Ceballos-Baumann, A.O. (2005). Transmodal sensorimotor networks during action observation in professional pianists. *J. Cogn. Neurosci*., 17, 282-293.

Haueisen J. and Knosche T. R., (2001). Involuntary motor activity in pianists evoked by music perception. *J. Cogn. Neurosci*., 13:6, 786-792

Hickok G., (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans*. J. Cogn. Neurosci*., 7, 1229-1243.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science,* 286, 2526–2528.

Keysers, C., Kohler, E., Umiltà, M.A., Nanettu, L. Fogassi, L., and Gallese, V. (2003). Audiovisual mirror neurone and action recognition. *Exp. Brain Res*., 153, 628-636.

Keller, P., Joint action in music performance. (2008) in F. Morganti, A. Carassa & G. Riva (Eds.), Enacting intersubjectivity: A cognitive and social perspective on the study of interactions (pp. 205–221). Amsterdam: IOS Press.

Keller, P., Knoblich, G., & Repp, B. (2007). Pianists duet better when they play with

themselves. *Consciousness and Cognition*, 16, 102–111.

Kilner, J. Vargas C., Duval S., Blackmore S.-J., Sirigu A., (2004). Motor activation prior to observation of a predicted movement. *Nat. Neurosci*. 7, 1299–1301.

Kilner L. M., Friston K. J. and Frith C. D., (2007). The mirror system: a Bayesian perspective. *NeuroReport,* Vol 18 No 6. 619-623.

Koelewijn T., van Schie H. T., Bekkering H., Oostenveld R., Jensen O., (2007). Motor-cortical beta oscillations are modulated by correctness of observed action. *Neuroimage,* 40, 767-775.

Koelsch, S., Friederici, A.D., (2003). Towards the neural basis of processing structure in music: comparative results of different neurophysiological investigation methods (EEG, MEG, fMRI). Ann. N. Y. Acad. Sci. 999, 15 – 27.

Koelsch, S., Frintz T., Schulze K., Alsop D., Schlaug G. (2005). Adults and children processing music: An fMRI study. *Neuroimage* 25, 1068–1076.

Kohler E, Keysers C, Umilta MA, Fogassi L, Gallese V, Rizzolatti G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science,* 297:846--848.

Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology (Paris)*, 102, 59–70.

Manthey S., Schubotz R. I., von Cramon D. Y., (2002). Premotor cortex in observing erroneous action: an fMRI study. *Cogn. Brain Res*., 15, 296-307.

Molnar-Szakacs I, Kaplan J, Greenfield PM, Iacoboni M. (2006). Observing complex action sequences: the role of the fronto-parietal mirror neuron system. *Neuroimage*, 33:923—935.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, 10, 70–76.

Slotnick, S.D., Moo, L.R., Segal, J.B., Hart J. (2003). Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Cognitive Brain Research* 17, 75-82.

Umiltà M. A., Kohler E., Gallese V., Fogassi L., Fadiga L., Keysers C. and Rizzolatti G. (2001). I Know What You Are Doing: A Neurophysiological Study. *Neuron*, Volume 31, Issue 1, Pages 155-165.

Van Elk M., van Schie H. T., Hunnius S., Vesper C., Bekkering H., (2008). You'll never crawl alone: neurophysiological evidence for experience-dependent motor resonance in infancy. *Neuroimage* 43, 808-814.

Verfaillie, K. and Daems, A. (2002). Representing and anticipating human actions in

vision. *Vis. Cogn*., 9, 217–232.

Vines B.W., Krumhansl C.L., Wanderley M.M., Levitin D.J., (2006). Cross-modal interactions in the perception of musical performance. *Cognition*, 101, 80-113.

Zatorre J. Z., Chen J. L., Penhune V. B., (2007). When the brain plays music: auditory–motor interactions in music perception and production*. Nature Reviews Neuroscience* 8, 547-558.

Figure captions

Figure 1: Examples of musical scores of the sequences (left part) and visual examples of sequences' last chord (right part). The figure shows a right hand performing (A) a congruent chord, (B) an auditory incongruent chord and (C) a motor incongruent chord. Letters indicate which fingers are used to play each chord. All conditions were presented without piano sound.

Figure 2: Performance rate (expressed in percentage) of non-musicians and pianists in audio and motor conditions. Panel A shows performance rate of expert pianists (left column) and non-musicians (right column) in the detection of auditory incongruent chords. Panel B shows performance rate of expert pianists (left column) and non-musicians (right column) in the detection of motor incongruent chords. Bars show chance level (50%). In the congruent condition, musicians performed at ca. 95%, and non-musicians performed at approximately chance (59%). The data are not shown here, as they provide us little information beyond that provided by performance in the two violation conditions.

Figure 3: Panel A shows difference in BOLD response for auditory violated trials in contrasts to congruent trials ($p<0.005$, $k>65$). Panel B shows difference in BOLD response for motor violated trials in contrasts to congruent trials ($p<0.005$, $k>65$).

Table 1: Performance rate of each subject in both the experimental conditions and the data from the post-experiment questionnaire are showed. Data represent respectively: percentage of correct trials in audio and motor conditions, number of years each player had played in the past, the age each player started to play the piano, the number of hours each player used to play (per week) the piano at the time of the experiment and the number of hours (per week) each player used to teach piano (if applicable).

Table 2: Brain regions significantly more active for auditory incongruencies than for congruent trials ($p<0.005$, $k> 65$). Maximum Z-scores, cluster extent (in voxels) and MNI co-ordinates are reported.

Table 3: Brain regions significantly more active for motor incongruencies than for congruent trials ($p<0.005$, $k> 65$). Maximum Z-scores, cluster extent (in voxels) and MNI co-ordinates are reported.

FIGURE 1

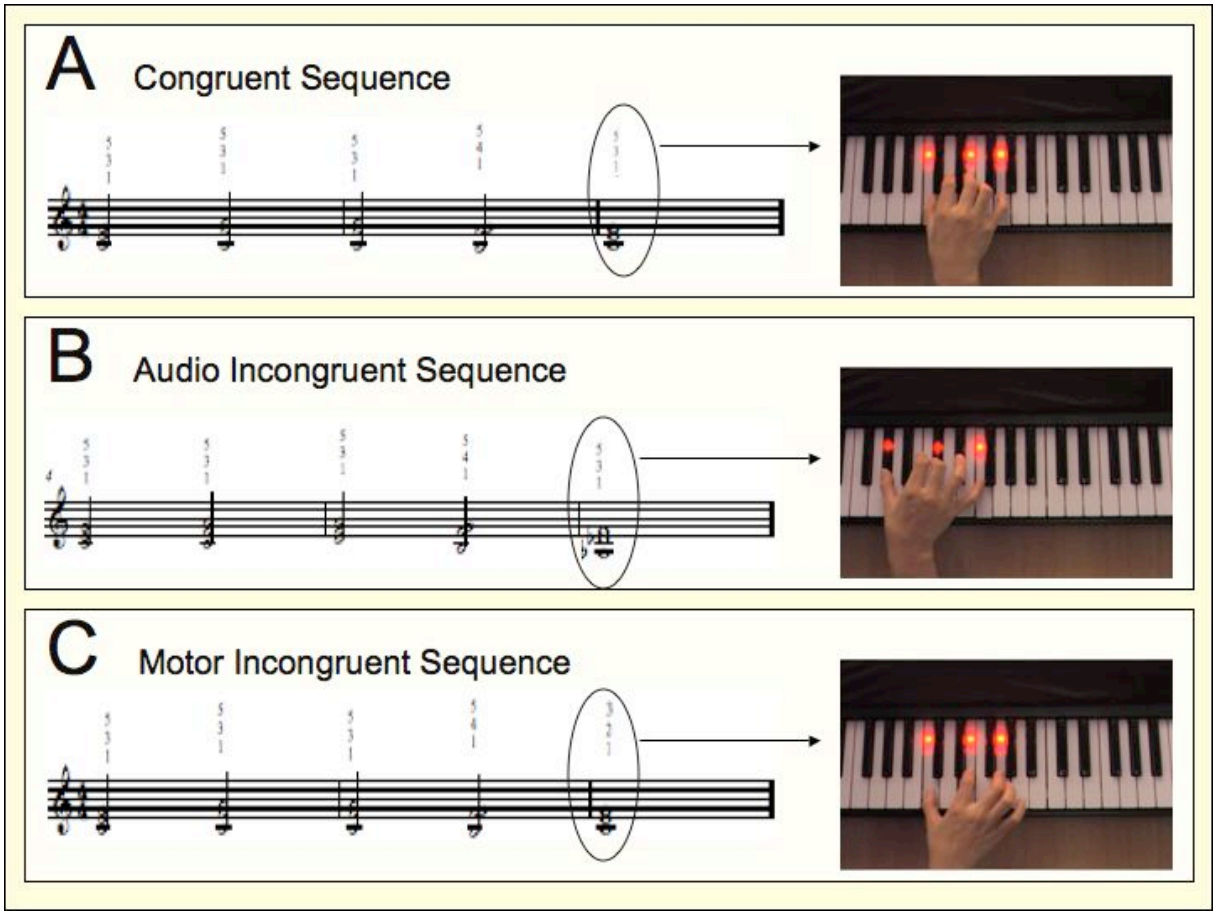

FIGURE 2

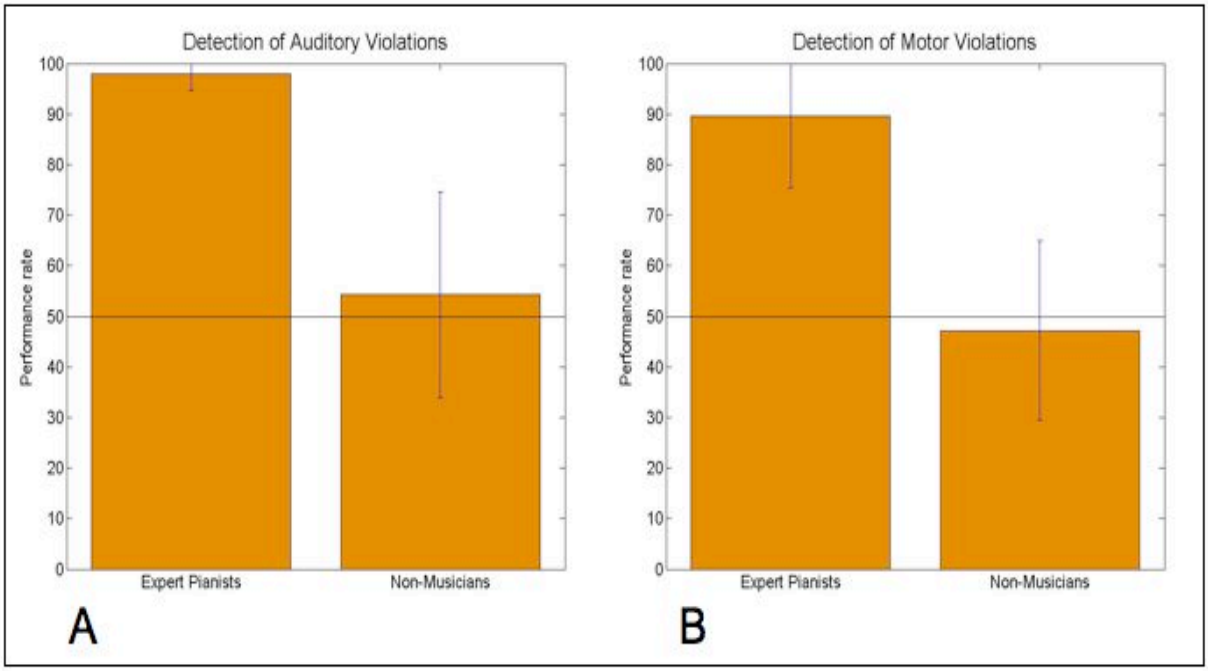

FIGURE 3

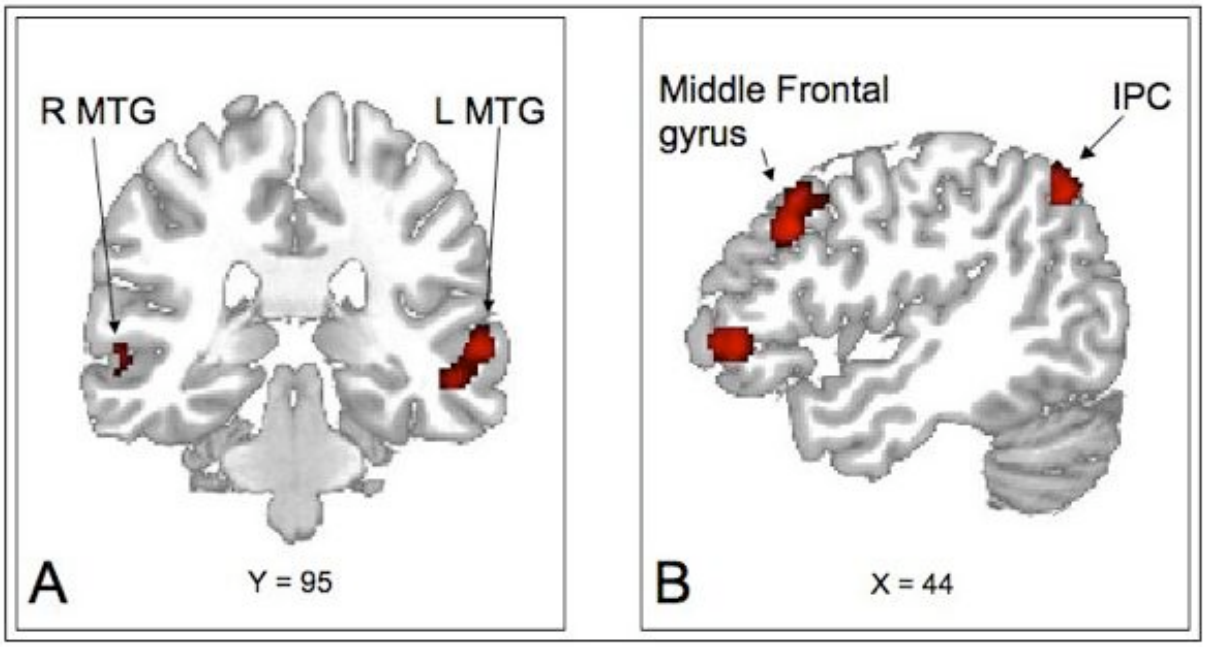

TABLE 1

Subject	Audio	Motor	Experience	Onset	Frequency	Teaching
01	100.0	56.0	12.0	9.0	17.0	1.5
02	100.0	92.0	13.0	5.0	35.0	0
03	94.0	91.0	16.0	5.0	24.0	1.5
04	100.0	100.0	14.0	6.0	20.0	1.5
05	100.0	83.0	15.0	9.0	28.0	0.5
06	97.0	100.0	12.0	5.0	30.0	0
07	100.0	100.0	21.0	4.0	32.0	2.5
08	97.0	77.0	13.0	10.0	30.0	2
09	91.0	100.0	21.0	7.0	35.0	18
10	100.0	97.0	15.0	11.0	30.0	0

TABLE 2

Brain Region	Zmax	Extent (voxels)	X	Y	Z
Left occipital cortex		2377			
Left Lingual Gyrus	3.99		-10	-90	-12
Left Cerebellum (Crus 1)	3.76		-46	-70	-20
Inferior Occipital Gyrus	3.72		-36	-80	-16
Left Middle Temporal Gyrus		145			
(Lateral-Superior) Left Middle Temporal Gyrus	3.66		-62	-32	2
(Medial-Inferior) Left Middle Temporal Gyrus	2.89		-56	-30	-6
Right Middle Temporal Gyrus		67			
(Medial-Inferior) Right Middle Temporal Gyrus	2.99		58	-38	-2
Right Parietal Cortex		680			
Right Superior Parietal Gyrus	3.64		20	-74	56
Right Superior Parietal Gyrus	3.57		30	-74	54
Right Superior Parietal Gyrus	3.52		46	-52	60
Right inferior occipital cortex		2879			
Inferior Occipital Gyrus	3.60		42	-72	-12
Right Cerebellum (Crus 1)	3.59		46	-56	-26
Middle Occipital Gyrus	3.54		34	-86	6
Left parietal cortex		488			
Left Superior Parietal Gyrus	3.43		-16	-70	64
Angular Gyrus	3.22		-38	-54	44
Left Superior Parietal Gyrus	2.92		-28	-66	62
Left middle frontal gyrus		143			
Middle Frontal Gyrus	3.07		-46	30	38
Middle Frontal Gyrus	3.06		-52	22	40
Left inferior frontal gyrus		88			
Pars Opercularis	2.94		-38	14	22
Pars Opercularis	2.79		-50	12	22
Right inferior parietal cortex		100			
Angular Gyrus	2.88		36	-76	28
Right middle frontal gyrus		125			
Right Middle Frontal Gyrus	2.86		34	24	36
Right Precentral Gyrus	2.86		56	12	42
Right Middle Frontal gyrus	2.76		48	22	48

TABLE 3

Brain Region	**Zmax**	**Extent (voxels)**	**X**	**Y**	**Z**
Left Superior Frontal Gyrus		70			
Left Superior Frontal Gyrus	3.75		-8	38	50
Left Middle Frontal Gyrus		134			
Left Middle Frontal Gyrus	3.72		-44	24	34
Left Middle Frontal Gyrus	3.42		-50	18	42
Left Inferior Frontal Gyrus		131			
Left Inferior Frontal Gyrus (p. angularis)	3.53		-50	38	-4
Anterior Middle Frontal Gyrus/ Inferior Frontal Solcus	2.87		-40	44	0
Left Inferior Frontal Gyrus (p. triangularis)	2.67		-52	34	6
Left Inferior parietal cortex		81			
Left Angular Gyrus	3.52		-50	-62	44
Left Angular Gyrus	3.06		-46	-62	52
Right inferior occipital gyrus		75			
Right Inferior Occipital gyrus	3.28		48	-72	-16
Inferior Occipital Gyrus	3.13		54	-66	-16