



OPEN

Audio-visuomotor processing in the Musician's brain: an ERP study on professional violinists and clarinetists

SUBJECT AREAS:
PERCEPTION
NEUROPHYSIOLOGYAlice Mado Proverbio¹, Marta Calbi^{1,2}, Mirella Manfredi^{1,3} & Alberto Zani⁴Received
2 April 2014Accepted
4 July 2014Published
29 July 2014Correspondence and
requests for materials
should be addressed to
P.A.M. (mado.
proverbio@unimib.it)

¹Milan Center for Neuroscience, University of Milano-Bicocca, Piazza dell'Ateneo Nuovo 1, 20126 Milan, Italy, ²Department of Neuroscience, University of Parma, Italy, ³University of California San Diego, La Jolla, California, ⁴National Research Council (CNR), Milan, Italy.

The temporal dynamics of brain activation during visual and auditory perception of congruent vs. incongruent musical video clips was investigated in 12 musicians from the Milan Conservatory of music and 12 controls. 368 videos of a clarinetist and a violinist playing the same score with their instruments were presented. The sounds were similar in pitch, intensity, rhythm and duration. To produce an audiovisual discrepancy, in half of the trials, the visual information was incongruent with the soundtrack in pitch. ERPs were recorded from 128 sites. Only in musicians for their own instruments was a N400-like negative deflection elicited due to the incongruent audiovisual information. SwLORETA applied to the N400 response identified the areas mediating multimodal motor processing: the prefrontal cortex, the right superior and middle temporal gyrus, the premotor cortex, the inferior frontal and inferior parietal areas, the EBA, somatosensory cortex, cerebellum and SMA. The data indicate the existence of audiomotor mirror neurons responding to incongruent visual and auditory information, thus suggesting that they may encode multimodal representations of musical gestures and sounds. These systems may underlie the ability to learn how to play a musical instrument.

The discovery of audiovisual mirror neurons in monkeys, a subgroup of premotor neurons that respond to the sounds of actions (e.g., peanut breaking) in addition to their visuomotor representation, suggest that there may be a similar cross-modal neural system in humans^{1,2}. We hypothesized that this neural system may be involved in learning how to play a musical instrument. Previous studies have shown that when playing an instrument (e.g., the piano), auditory feedback is naturally involved in each of player's movements, leading to a close coupling between perception and action^{3,4}. In a recent study, Lahav et al.⁵ investigated how the mirror neuron system responds to actions and sounds of well-known melodies compared to new piano pieces. The results revealed that music the subject knew how to play was strongly associated with the corresponding elements of the individual's motor repertoire and activated an audiomotor network in the human brain. However, the whole-brain functional mechanism underlying such an "action-listening" system is not fully understood.

The advanced study of music involves intense stimulation of sensory, motor and multimodal neuronal circuits for many hours per day over several years. Very experienced musicians are capable of otherwise unthinkable capacities, such as recognizing if a violinist is playing a slightly flat or sharp note solely based on the position of their hand on the fingerboard. These capabilities result from a long training, during which imitative processes play a crucial role.

One of the most striking manifestations of the multimodal audiovisual coding of information is the McGurk effect⁶, which is a linguistic phenomenon observed during audiovisual incongruence. For example, when the auditory component of one syllable (e.g., \ba\) is paired with the visual component of another syllable (e.g., \ga\), the perception of a third syllable (e.g., \da\) is induced, thus suggesting a multimodal processing of information. Calvert and colleagues⁷ investigated the neural mechanisms subserving the McGurk effect in an fMRI study in which participants were exposed to various fragments of semantically congruent and incongruent audio-visual speech and to each sensory modality in isolation. The results showed an increase in the activity of the superior temporal sulcus (STS) for the multimodal condition compared to the unimodal condition. To correlate brain activation with the level of integration of audiovisual information, Jones and Callan⁸ developed an experimental paradigm based on phoneme categorization in which the synchrony between audio and video was systematically manipulated. fMRI revealed a greater parietal activation at the right supramarginal gyrus and the left inferior



parietal lobule during incongruent stimulation compared to congruent stimulation. Although fMRI can be used to identify the regions involved in audiovisual multisensory integration, neurophysiological signals such as EEG/MEG, especially in Mismatch Negativity (MMN) paradigms, can provide information regarding the timing of this activation, especially if the timing involves qualitative changes in the primary auditory cortex, and whether integration occurs at later cognitive levels. MMN is a response of the brain that is generated primarily in the auditory cortex. The amplitude of the MMN response depends on the degree of variations/changes in the expected auditory percept, thus reflecting the cortical representation of auditory-based information⁹. Sams and collaborators¹⁰ used a MMN paradigm to study the McGurk effect and found that deviant stimuli elicited a MMN generated at level of primary auditory cortex, suggesting that visual speech processing can affect the activity of the auditory cortex^{11,12} at the earliest stage.

Besle and coworkers¹³ recorded intracranial ERPs evoked by syllables presented in three different conditions (only visual, only auditory and multimodal) from depth electrodes implanted in the temporal lobe of epileptic patients. They found that lip movements activated secondary auditory areas very shortly (≈ 10 ms) after the activation of the visual motion area MT/V5. After this putative feed forward visual activation of the auditory cortex, audiovisual interactions took place in the secondary auditory cortex, from 30 ms after the sound onset and prior to any activity in the polymodal areas. Finally, in a MEG study, Mottonen et al.¹⁴ found that viewing the articulatory movements of a speaker emphasizes the activity in the left mouth primary somatosensory (SI) cortex of the listener. Interestingly, this effect was not seen in the homologous right SI, or even in the SI corresponding to the hands in both hemispheres of the listener. Therefore, the authors concluded that visual processing of speech activates the corresponding areas of the SI in a specific somatotopic manner.

Similarly to audiovisual processing of phonetic information, multimodal processing may play a crucial role in audiomotor music learning. In this regard, MMN can be a valuable tool for investigating multimodal integration and plasticity in musical training^{15,16}. For example, Pantev et al.¹⁵ trained a group of non-musicians, the sensorimotor-auditory group (SA), to play a musical sequence on the piano while a second group, the auditory group (A), actively listened to and made judgments about the correctness of the music. The training-induced cortical plasticity effect was assessed via magnetoencephalography (MEG) by recording musically elicited (MMN) before and after the training. The SA group showed a significant enlargement of MMN after training compared to the A group, reflecting a greater enhancement of musical representations in the auditory cortex after sensorimotor-auditory training compared to auditory training alone. In another MMN study¹⁶, it was found that the cortical representations for notes of different timbre (violin and trumpet) were enhanced in violinists and trumpeters, preferentially for the timbre of the instrument on which the musician was trained, and especially when both parts used to play the instruments were stimulated (cross-modal plasticity). For example, when the lips of trumpet players were stimulated touching the mouthpiece of their instrument at the same time as a trumpet tone, activation in the somatosensory cortex increased more than the sum of the somatosensory activation increases for lip touch and trumpet audio stimulation administered separately.

In an fMRI study¹⁷, it was investigated how pianists are able to encode the association between the visual display of a sequence of key pressing in a silent movie and the corresponding sounds, thus enabling them to recognize which piece was being played. In this study, the temporal planum was found to be heavily involved in multimodal coding. The most experienced pianists exhibited a bilateral activation of the premotor cortex, the inferior frontal cortex, the parietal cortex and the SMA, similar to the findings of Schuboz and von Cramon¹⁸.

McIntosh and colleagues¹⁹ examined the effect of audiovisual learning in a crossmodal condition with positron emission tomography (PET). In this study, participants learned that an auditory stimulus systematically signaled a visual event. Once learned, activation of the left dorsal occipital cortex (increased regional CBF) was observed when the auditory stimulus was presented alone. Functional connectivity analysis between the occipital area and the rest of the brain revealed a pattern of covariation with four dominant brain areas that may have mediated this activation: the prefrontal, premotor, superior temporal, and contralateral occipital cortices.

Notwithstanding previous studies, knowledge regarding the neural bases of music learning is still quite scarce. The present work aimed to investigate the timing of activation and the role of multisensory audiomotor and visuomotor areas in the coding of musical sounds associated with musical gestures in experienced musicians. We sought to record the electromagnetic activity of systems similar to multimodal neurons that code both phonological sounds and lip movements in language production/perception. In addition to source reconstruction neuroimaging data (provided by swLORETA) we aimed to gain precious temporal information about synchronized bioelectrical activity during perception of a music execution, at the millisecond resolution.

Undergraduate, master students and faculty professors at *Verdi Conservatory* in Milan were tested under conditions incorporating a violin or clarinet, depending on the instrument played by the subject. Musicians were subjected to stimulation by presenting movie clips in which a colleague executed sequences of single or paired notes. We filmed 2 musicians who were playing either the violin or the clarinet.

Half of the clips were then manipulated such that, although perfectly synchronized in time, the videos' soundtrack did not correspond with the note/s actually played (incongruent condition). For these clips, we hypothesized that the mismatch between visual and auditory information would stimulate multimodal neurons that encode the audio/visuomotor properties of musical gestures; indeed, expert musicians have acquired through years of practice the ability to automatically determine whether a given sound corresponds with the observed position of the fingers on the fingerboard or set of keys. We predicted that the audio-video inconsistency would be clearly recognizable only by musicians skilled in that specific instrument (i.e., in violinists for the violin, and in clarinetists for the clarinet), provided that musicians were unskilled at using the other musical instrument. Before testing, stimuli were validated by a conspicuous group of independent judges (recruited at Milan Conservatory "Giuseppe Verdi") that established how easily the soundtrack inconsistency was recognizable.

Two different musical instruments were considered in this study for multiple reasons. First, i) this design provides the opportunity to compare the skilled vs. unskilled audiomotor mechanisms within a musician's brain, as there are many known differences between musicians' and non-musicians' brains at both the cortical and sub-cortical level²⁰. It is well known, for example, that musical training since infancy results in changes in brain connectivity, volume, and functioning²¹, in particular in motor performance (basal ganglia, cerebellum, motor and premotor cortices), visuomotor transformation (the superior parietal cortex)^{22,23}, inter-hemispheric callosal exchanges²⁴, auditory analysis^{25,26} and the notation reading (Visual Word Form Area, VWFA)²⁷ are concerned (see Kraus & Chandrasekaran²⁸ for a review). Furthermore, several studies have compared musicians with non-musicians, highlighting a number of structural and functional differences in the sensorimotor cortex^{22,23,29,30} and areas devoted to multi-sensory integration^{22,23,31,32}. In addition, neural plasticity seems to be very sensitive to the conditions during which multisensory learning occurs. For example, it was found that violinists have a greater cortical representation of the left compared to the right hand²⁹, trumpeters exhibit a stronger interaction between the auditory and somatosensory inputs relative

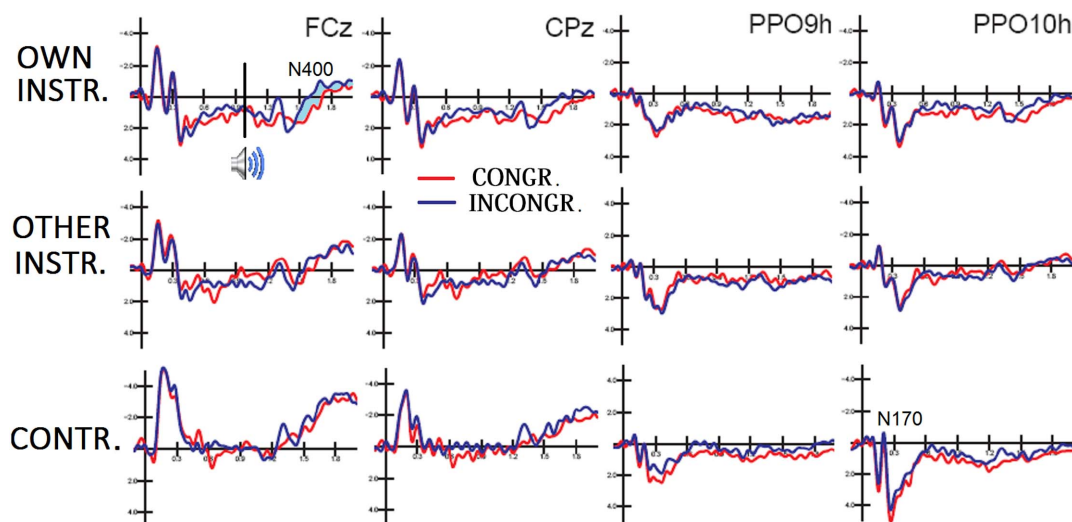


Figure 1 | Grand-average ERP waveforms recorded from the midline fronto-central (FCz), the centro-parietal (Cpz), and the left and right occipito-temporal (PPO9h, PPO10h) sites as a function of group and stimulus audiovisual congruence. No effect of condition (congruent vs. incongruent) is visible in controls and in musicians for the unfamiliar instrument.

to the lip area³³, and professional pianists show a greater activation in the supplementary motor area (SMA) of the cortex and the dorso-lateral premotor cortex³⁴ compared to controls. These neuroplastic changes concern not only the gray matter but also the white fibers³⁵ and their myelination³⁶.

Moreover, ii) we aimed to investigate the general mechanisms of neural plasticity, independent of the specific musical instrument played (strings vs. woods) and the muscle groups involved (mouth, lips, left hand, right hand, etc.)

In the ERP study, brain activity during audiovisual perception of congruent vs. incongruent sound/gesture movie clips was recorded from professional musicians who were graduates of Milan Conservatory “Giuseppe Verdi” and from age-matched university students (controls) while listening and watching violin and clarinet executions. Their task was to discriminate a 1-note vs. 2-note execution by pressing one out of two buttons. The task was devised to be feasible for both naïve subjects and experts and to allow automatic processing of audiovisual information in both groups, according to their musical skills. EEG was recorded from musicians and controls to record the bioelectrical activity corresponding to the detection of an audiovisual incongruity. In paradigms where a series of standard stimuli followed by the deviant stimuli were presented, the incongruity typically elicit a visual Mismatch Negativity (vMMN)^{37,38}. In this study, we expected to find an anterior N400-like negative deflection sharing some similarities with a vMMN, but occurring later due to the dynamic nature of the stimulus (movies lasting 3 seconds). Previous studies have identified anterior N400 to incongruent gestures on action processing when a violation was presented, such as a symbolic hand gesture³⁹, a sport action⁴⁰, goal directed behavior^{41,42}, affective body language⁴³, or an action-object interaction^{44,45}. We expected to find a significantly smaller or absent N400 in the musicians’ brains in response to violations relative to the instrument which the subject did not play and a lack of the response the naïve subjects’ brains.

Results

Behavioral data. ANOVA performed on accuracy data (incorrect categorizations) revealed no effect of the group on the error percentage, which was below 2% ($F_{1,30} = 0.1295$; $p = 0.72$), or hits percentage ($F_{1,3} = 0.3879$; $p = 0.538$).

ANOVA performed on response times indicated ($F_{1,22} = 6.7234$; $p < 0.017$) longer RTs ($p < 0.02$) in musicians (2840 ms, 1840 ms

post-sound latency, $SE = 81.5$) compared to controls (2614 ms, 1641 ms post-sound latency, $SE = 81.5$).

ERP data. Figure 1 shows the grand-average ERPs recorded in response to congruent and incongruent stimulation, independent of the musical instrument but considering participants’ expertise, in musicians and controls (instruments were collapsed). An N400-like response at anterior sites was observed in musicians under only conditions incorporating their own musical instrument, which was characterized by an increased negativity for incongruent soundtracks compared to congruous soundtracks between the post-sound 500 to 1000 ms time window.

N170 component. ANOVA performed on the N170 latency values revealed significance of the hemisphere factor ($F_{1,22} = 11.36$; $p < 0.0028$), with faster N170 latencies recorded over the LH (173 ms, $SE = 2.1$) compared to the RH (180 ms, $SE = 2.5$). Interestingly, the N170 latency was also affected by the group factor ($F_{1,22} = 9.2$; $p < 0.0062$). Post-hoc comparisons indicated faster latencies of the N170 response in Musicians when using his/her Own instrument (164 ms, $SE = 3.9$) compared with the Other Instrument ($p < 0.05$; 176 ms, $SE = 4.2$) and compared with controls ($p < 0.008$; 183 ms, $SE = 4.1$).

N400. ANOVA computed on the mean amplitude of the negativity recorded from 500–1000 ms post-sound stimulation revealed a greater amplitude at anterior site (Fcz, $-1.89 \mu V$, $SE = 0.42$) compared with the central ($p < 0.01$; Cz $-1.78 \mu V$, $SE = 0.44$) and centroparietal ($p < 0.001$; Cpz $-1.25 \mu V$, $SE = 0.42$) sites, as indicated by a significant electrode factor ($F_{2,44} = 7.78$; $p < 0.01$) and post-hoc comparisons. ANOVA also yielded a significant Condition effect ($F_{1,22} = 7.35$, $p < 0.02$) corresponding to a greater N400 amplitude in response to Incongruent videos ($-1.84 \mu V$, $SE = 0.42$) compared to Congruent videos ($-1.44 \mu V$, $SE = 0.42$). A significant Electrode \times Group interaction ($F_{2,44} = 3.25$; $p < 0.05$) revealed larger N400 responses at the anterior site in the control group (Fcz, $-2.79 \mu V$, $SE = 0.60$; Cz, $-2.25 \mu V$, $SE = 0.62$; CPz, $-1.86 \mu V$, $SE = 0.60$) compared to the musician group (Fcz, $-0.99 \mu V$, $SE = 0.60$; Cz, $-1.31 \mu V$, $SE = 0.62$; CPz, $-0.64 \mu V$, $SE = 0.60$), which was confirmed by post-hoc tests ($p < 0.006$). However, the N400 amplitude was strongly modulated by Condition only in musicians and in scenarios that incorporated their own musical instrument (see ERP waveform of Fig. 2), as revealed by the significant Instrument \times Condition \times Group interaction ($F_{1,22} = 11.73$, $p < 0.003$). Post-



hoc comparisons indicated a significant ($p < 0.007$) N400 enhancement in response to the Own instrument for Incongruent videos ($-0.86 \mu\text{V}$, $\text{SE} = 0.74$) compared with Congruent videos ($-0.23 \mu\text{V}$, $\text{SE} = 0.76$). Moreover, no significant differences ($p = 0.8$) were observed in musicians in response to the Other instrument for Incongruent ($-1.53 \mu\text{V}$, $\text{SE} = 0.67$) vs. Congruent videos ($-1.31 \mu\text{V}$, $\text{SE} = 0.63$). For the control group, no differences ($p = 0.99$) emerged between the Congruent and Incongruent contrast for either instruments. Finally, the ANOVA revealed a significant Instrument \times Condition \times Electrode \times Group interaction ($F_{2,44} = 3.89$; $p < 0.03$), revealing additional significant group differences in the responses to incongruent vs. congruent stimuli at the anterior site compared with the central and posterior sites, as shown in Figure 3.

To investigate the neural generators of violation-related negativity in musicians (Own instrument), a swLORETA inverse solution was applied to the difference wave obtained by subtracting ERPs recorded during Congruent stimulation from ERPs recorded during Incongruent stimulation in the 500–1000 (post-sound) time window (see Table 1 for a list of relevant sources). SwLORETA revealed a complex network of areas with different functional properties active during the synchronized mismatch N400 response to audiovisual incongruence. The strongest sources of activation were the anterior site and the cognitive discrepancy perception (Left and right BA10) areas, as shown in Fig. 4 (Bottom, rightmost axial section). Other important sources were the right temporal cortex (superior temporal gyrus, or BA38, and the middle temporal gyrus, or BA21), regions belonging to the “human mirror neuron system (MNS)” (i.e., the premotor cortex, or BA6, inferior frontal area, or BA44, and the inferior parietal lobule, or BA40), areas devoted to body or action

representations (the extrastriate body area, (EBA), or BA37) and somatosensory processing (BA7), and motor regions, such as the cerebellum and the supplementary motor area (SMA) (see the rightmost axial section in the top row of Fig. 4).

Discussion

In this study, the effects of prolonged and intense musical training on the audiovisual mirror mechanism was observed by investigating the temporal dynamics of brain activation during audiovisual perception of congruent vs. incongruent sound-gesture movie clips in musicians and naïve age-matched subjects.

To ensure the subject’s attention was focused on stimulation, we instructed participants to respond as quickly as possible to stimuli and to decide whether the musicians in the movie had played one or two tones. No effect of audiovisual match was observed on behavioral performance. Musicians tended to be a bit slower than controls, most likely because they have a more advanced musical understanding.

ERPs revealed that experts exhibited an earlier N170 latency to visual stimulation. The view of a musician playing was processed much earlier if the instrument was their own compared to an unfamiliar instrument and the response was overall faster in musicians than in controls, indicating an effect of visual familiarity for the musical instrument.

For this reason, two different instruments were considered in this study, and the reciprocal effect of expertise was investigated within a musicians’ brains (whether skilled or not) compared to the brains of non-musicians. A negative drift is visible in the ERP waveforms shown in Fig. 2 at the anterior electrode sites only, which started at approximately 1500 ms post-stimulus in the musicians’ brains but

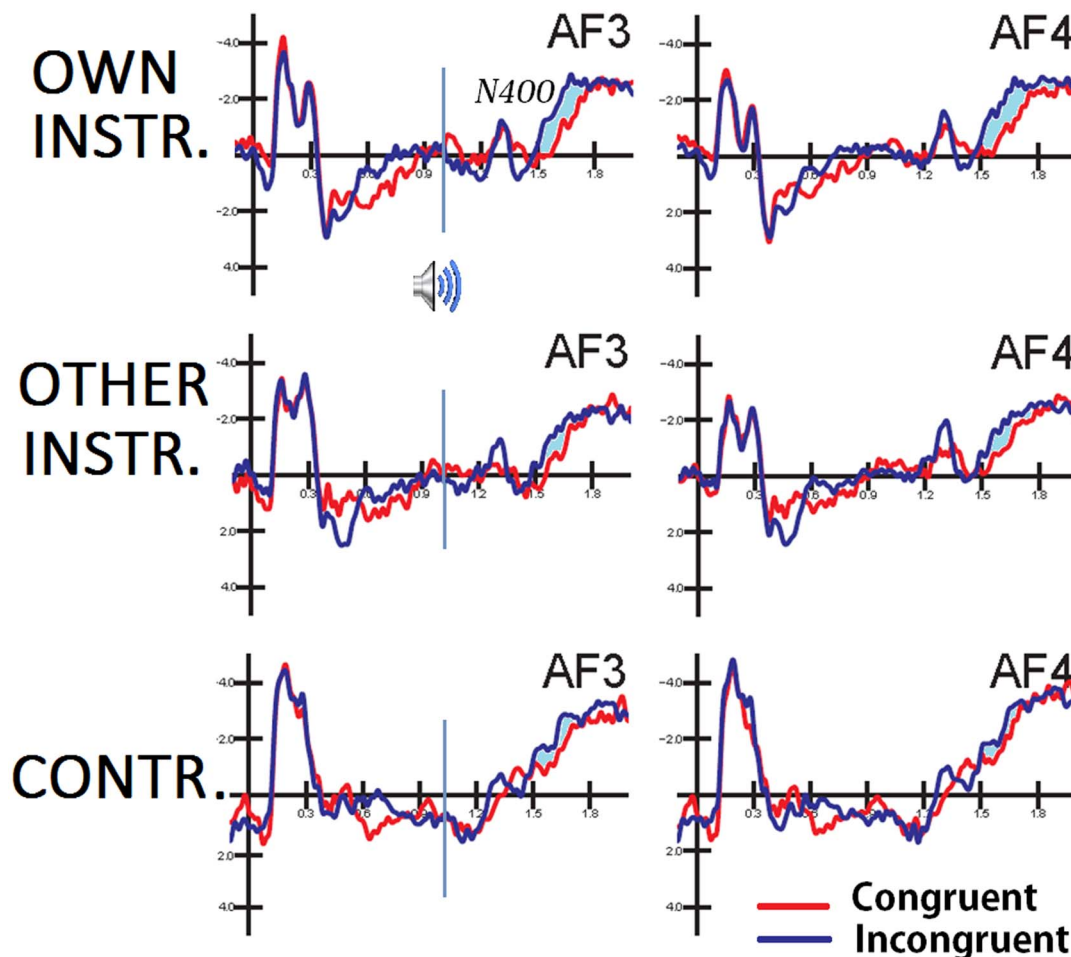


Figure 2 | Grand-average ERP waveforms recorded at the left and right anterior frontal sites as a function of group and stimulus congruence.

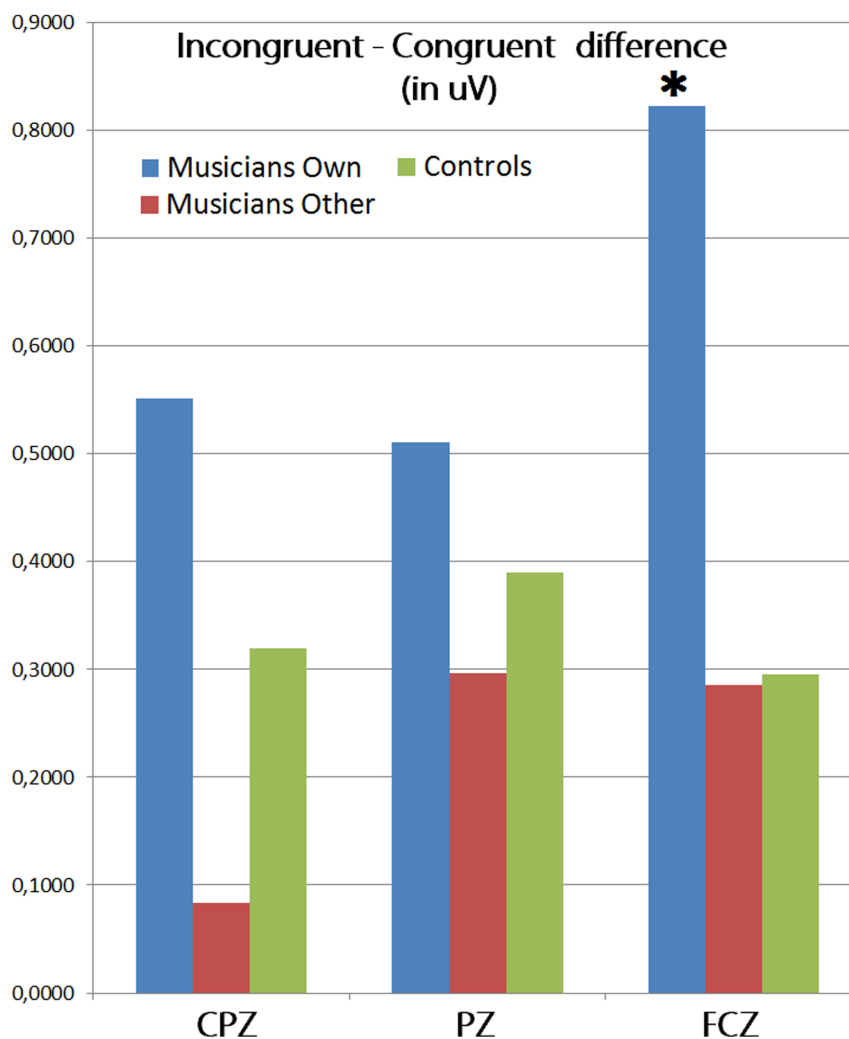


Figure 3 | Mean amplitude (μV) of the incongruent-congruent differential N400 response recorded in musicians and controls at the anterior, central and centroparietal sites. The only significant task-related effect was found in musicians for their own instrument at frontal sites.

Table 1 | Talairach coordinates (in mm) corresponding to the intracortical generators that explain the surface voltage recorded during the 1500–2000 ms time window in response to incongruent vs. congruent clips in the musicians' brain during scenarios incorporating their own musical instrument. Magn. = Magnitude in nAm; H = hemisphere, BA = Brodmann areas

| Incongruent-Congruent (500-1000 ms) - Power RMS = 51.8 | | | | | | | | |
|--|-------|-------|-------|------|--------|----------------------|----|-----------------------|
| Magn. | T-x | T-y | T-z | Hem. | Lobe | Gyrus | BA | Function |
| 11.38 | -8.5 | 64.4 | 16.8 | L | Front | Sup. Frontal | 10 | Cognitive Discrepancy |
| 11.05 | -28.5 | 53.4 | 24.8 | L | Front | Sup. Frontal | 10 | |
| 9.46 | 40.9 | 55.3 | 7 | R | Front | Middle Frontal | 10 | Sound processing |
| 8.08 | 50.8 | -0.6 | -28.2 | R | Temp | Middle Temporal | 21 | |
| 7.59 | 31 | 9.1 | -27.5 | R | Temp | Sup. Temporal | 38 | Body/face processing |
| 7.23 | 60.6 | -55 | -17.6 | R | Occip | Fusiform Gyrus | 37 | |
| 7.02 | 50.8 | -33.7 | -23.6 | R | Temp | Fusiform Gyrus | 20 | Object processing |
| 6.82 | -38.5 | -8 | -28.9 | L | Temp | Middle Temporal | 21 | |
| 6.77 | -58.5 | -8.7 | -21.5 | L | Temp | Inferior Temporal | 20 | Motor coordination |
| 6.07 | 40.9 | -75.2 | -19.1 | R | | Cerebellum | | |
| 7.08 | -18.5 | -8 | -28.9 | L | Limbic | Uncus | 36 | Affective reaction |
| 4.76 | -58.5 | -30.4 | 34.9 | L | Pariet | Inf. Parietal Lobule | 40 | Action |
| 4.31 | 21.2 | -91.3 | 29.7 | R | Occip | Cuneus | 19 | Visual sensory |
| 3.9 | 1.5 | -85.4 | -19.8 | R | | Cerebellum | | Motor coordination |
| 3.88 | -8.5 | -1.1 | 65 | L | F | Sup. Frontal | 6 | SMA |
| 3.55 | -58.5 | 14.3 | 12.5 | L | F | Inf. Frontal | 44 | Mirror neurons |
| 2.94 | -18.5 | -63.8 | 59 | L | P | Sup. Parietal Lobule | 7 | Somatosensory |
| 2.37 | 11.3 | 29.5 | 58.7 | R | F | Sup. Frontal | 6 | SMA |

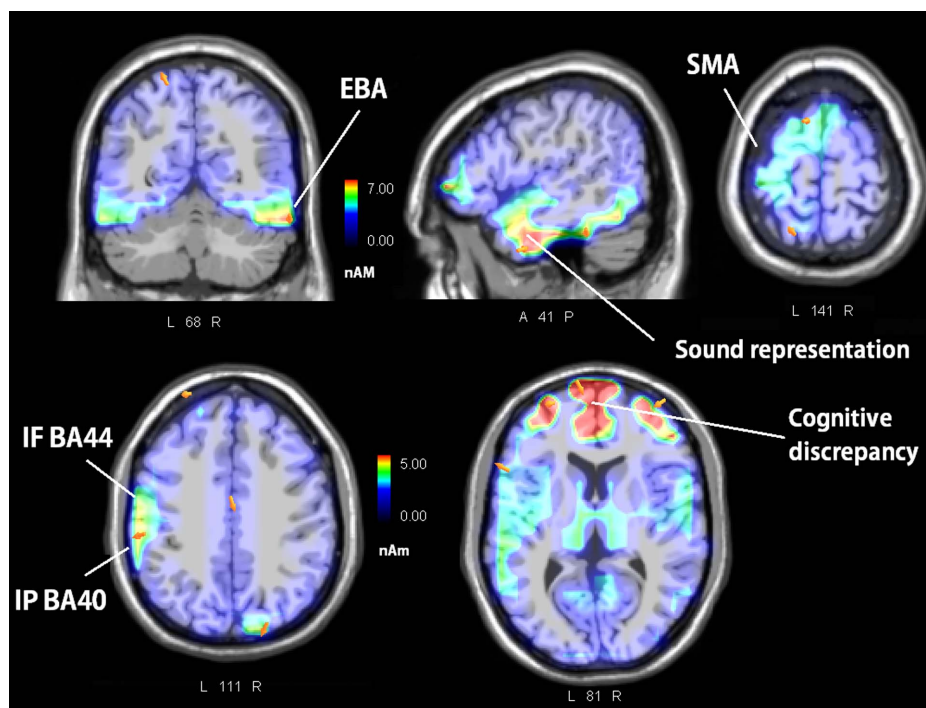


Figure 4 | Coronal, sagittal and axial views of the N400 active sources for the processing of musical audiovisual incongruence according to swLORETA analysis during 500–1000 ms post-sound start. The various colors represent differences in the magnitude of the electromagnetic signal (nAm). The electromagnetic dipoles are shown as arrows and indicate the position, orientation and magnitude of the dipole modeling solution applied to the ERP waveform in the specific time window. L = left; R = right; numbers refer to the displayed brain slice in the MRI imaging plane.

several hundreds of ms earlier in the naïve subjects' brains. This increase in negativity at the anterior sites (for all groups) possibly represents a contingent negative variation (CNV) linked to motor programming that precedes the upcoming button press response. It could be hypothesized that the CNV started earlier in the control group than the musician group, as control subjects' RTs were 200 ms faster than those of musicians. In addition to the CNV being initiated earlier in controls, it was larger than that of musicians at 500–1000 ms post-sound latency (N400 response). Importantly, the ERP responses were not modulated in amplitude by the stimulus content, as shown by the statistical analysis performed on N400 amplitudes.

Analyses of post-sound-start-related ERPs (occurring after 1000 ms post-stimulus) indicated that the automatic perception of soundtrack incongruence elicited an enlarged N400 response at the anterior frontal sites in the 500–1000 ms time window only in musicians' brains and for scenarios incorporating their own musical instrument.

The fact that the first executed note lasted 1 second in the 2-note condition and lasted 2 seconds in the minimum condition suggests that the early occurrence of the N400, before the sound was perceived, was initiated by the performer. These data suggest an automatic processing of audiovisual information. Considering that the task was implicit because participants were asked to determine the number of notes while ignoring other types of information, these findings also support the hypothesis that the N400 may share some similarities with the vMMN, which is generated in the absence of voluntary attention mechanisms^{46,47}. However, the possibility that the audiovisual incongruence attracted the attention of musicians after its automatic detection cannot be ruled out. This phenomenon occurred only in musicians during scenarios incorporating their own instrument and was not observed for the other groups or conditions. A similar modulation of the vMMN for audiovisual incongruent processing has been previously identified for the linguistic McGurk effect^{10–12}.

The presence of an anterior negativity in response to a visual incongruence was also reported in a study that compared the proces-

sing of congruent and expected vs. incoherent and meaningless behavior (e.g., in tool manipulation or goal-directed actions). In previous ERP studies, perception of the latter type of scenes elicited an anterior N400 response, reflecting a difficulty to integrate incoming visual information with sensorimotor-related knowledge⁴⁰. Additionally, in a recent study, Proverbio et al.⁴¹ showed that the perception of incorrect basketball actions (compared to correct actions) elicited an enlarged N400 response at anterior sites in the 450–530 ms time window in professional players, suggesting that action coding was automatically performed and that skilled players detected the violation of basketball rules.

In line with previous reports, in the present study, we found an enlarged N400 in response to incorrect sound-gesture pairs only in musicians, revealing that only skilled brains were able to recognize an action-sound violation. These results can be considered the electrophysiological evidence of a “hearing–doing” system⁵, which is related to the acquisition of nonverbal long-lasting action-sound associations. A swLORETA inverse solution was applied to the Incongruent–Congruent difference ERP waves (Own instrument condition) in the musician groups. This analysis revealed the premotor cortex (BA6), the supplementary motor area (BA6), the inferior parietal lobule (BA40), which has been shown to code transitive motor acts and meaningful behavioral chains (e.g., brushing teeth or flipping a coin), and the inferior frontal area (BA44) as the strongest foci of activations. Previous studies^{48–50} have shown the role of these regions in action recognition and understanding (involving the MNS). Indeed, the MNS is not only connected with the motor actions but also with linguistic “gesture” comprehension and vocalization. Several transcranial magnetic stimulation (TMS) studies have shown an enhancement of motor evoked potentials over the left primary motor cortex during both the viewing of speech⁵¹ and listening⁵².

Our findings support the data reported by Lahav et al.⁵ regarding the role of the MNS in audiomotor recognition of newly acquired actions (trained- vs. untrained-music). In addition, in our study, swLORETA showed the activation of the superior temporal gyrus



Score by M° Loguercio (Two Notes)

Figure 5 | An excerpt of the musical score played by the musicians to create the audiovisual stimuli.

(BA38) during sound perception. This piece of data suggests that the visual presentation of musical gestures activates the cortical representation of the corresponding sound only in skilled musicians' brains. In addition to being an auditory area, the STS is strongly interconnected with the fronto/parietal MNS⁵³. Overall, our data further confirm previous evidence of increased motor excitability⁵⁴ and premotor activity⁵⁵ in subjects listening to a familiar musical piece, thus suggesting the existence of a multimodal audiomotor representation of musical gestures. Furthermore, our findings share some similarities with previous studies that have shown a resonance in the mirror system of skilled basketball players⁴⁰ or dancers⁵⁶ during observation of a familiar motor repertoire or movements from their own dance style but not from other styles. Indeed, in the musician brain, a N400 was not found in response to audiovisual incongruence for the unfamiliar instrument (that is, the violin for clarinetists, and the clarinet for violinist).

Additionally, swLORETA revealed a focus of activation in the lateral occipital area, also known as the extrastriate body area (EBA, BA37), which is involved in the visual perception of body

parts, and of the right fusiform gyrus (BA37), a region that includes both the fusiform face area (FFA)⁵⁷ and the fusiform body area⁵⁸, which are selectively activated by human faces and bodies, respectively. These activations are most likely linked to the processing of musicians' fingers, hands, arms, faces and mouths/lips. The activation of cognitive-related brain areas, such as the superior and middle frontal gyrus (BA10), to stimulus discrepancy may be related to an involuntary attention orientation to visual/sound discrepancies at the pre-perceptual level^{59,60}. This finding supports the hypothesis that the detection signal generated by the violation within the auditory cortex is able to automatically trigger the orienting of attention at frontal fronto-polar level^{46,61,62}.

In conclusion, the results of the present study show a highly specialized cortical network in the skilled musician's brain that codes the relationship between gestures (both their visual and sensorimotor representation) and the corresponding sounds that are produced, as a result of musical learning. This information is very accurately coded and is instrument-specific, as indicated by the lack of an N400 in musicians' brains in scenarios incorporating the unfamiliar



Figure 6 | Frames taken from the video clips relative to clarinet and violin instruments. For the clarinetist, the lateral view allowed vision of the tonehole (above the musician's left thumb); for the violinist, the seated position allowed a clear view of finger on the fingerboard.

(Other) instrument. This finding bears some resemblance to the MEG data reported by Mottonen et al.¹⁴, which demonstrated that viewing the articulatory movements of a speaker specifically activates the left SI mouth cortex of the listener, resonating in a very precise manner from the sensory/motor point of view. Notwithstanding the robustness and soundness of our source localization data, it should be considered that some limitations in spatial resolution are intrinsic to EEG techniques because the bioelectrical signal becomes distorted while travelling through the various cerebral tissues and because EEG sensors can pick up only post-synaptic potentials coming from neurons whose apical dendrites are oriented perpendicularly to the recording surface⁶³. For these reasons, the convergence of interdisciplinary methodologies, such as fMRI data reported by Lahav et al.⁵ and MEG data reported by Mottonen et al.¹⁴, are particularly important for the study of audiomotor and visuomotor mirror neurons.

Methods

Participants. Thirty-two right-handed participants (8 males and 24 females) were recruited for the ERP recording session. The musician group included 9 professional violinists (3 males) and 8 professional clarinetists (3 males). The control group included 15 age-matched psychology students (2 males). Eight participants were discarded from ERP averaging due to technical problems during EEG recording (3 controls); therefore, 12 musicians were compared with 12 controls in total. The control subjects had a null experience with the violin and clarinet and never received specific musical education. The mean ages of violinists, clarinetists, and controls were 26 years (SD = 3.54), 23 years (SD = 3.03) and 23.5 years (SD = 2.50), respectively. The mean age of acquisition (AoA) of musical abilities (playing an instrument) was 7 years (SD = 2.64) for violinists and 10 years for clarinetists (SD = 2.43). The age ranges were 22–32 years for violinists and 19–28 years for clarinetists. The AoA ranges were 4–11 years for violinists and 7–13 years for clarinetists.

All participants had a normal or corrected vision with right eye dominance. They were strictly right-handed as assessed by the Edinburgh Inventory and reported no history of neurological illness or drug abuse. Experiments were conducted with the understanding and written consent of each participant according to the Declaration

of Helsinki (BMJ 1991; 302: 1194), with approval from the Ethical Committee of the Italian National Research Council (CNR) and in compliance with APA ethical standards for the treatment of human volunteers (1992, American Psychological Association).

Stimuli and procedure. A musical score of 200 measures was created (in 4/4 tempo), featuring 84 single note measures (1 minim) and 116 double note measures (2 semiminims). Single notes were never repeated and covered the common extension of the 2 instruments (violin and clarinet). Each combination of the two sounds was also absolutely unique. Stimulus material was obtained by videotaping a clarinetist and a violinist performing the score. Fig. 5 shows an excerpt from the score written by one of the violin teachers at Conservatory. Music was executed non-legato, and moderately vibrato on the violin (metronome = BPM 60) for approximately 2 seconds of sound stimulation for each musical beat. The two videos, one for each instrument, were subsequently segmented into 200 movie clips for each instrument (as an example of stimuli, see the initial frame of 2 clips relative to the two musical instruments in Fig. 6). Each clip lasted 3 seconds: during the first second the musician readied himself but did not play, and during the second 2 sec the tones were played. The average luminance of the violin and clarinet clips was measured using a *Minolta luminance meter*, and luminance values underwent an ANOVA to confirm equilibrium between the two stimulus classes (violin = 15.75 cd/m²; clarinet = 15.57 cd/m²). Audio sound values were normalized to -16 dB using the *Sony Sound Forge 9.0* software, by setting a fixed value of the root mean square (RMS) of a sound corresponding to the perceived intensity recorded at intervals of 50 ms. To obtain an audiovisual incongruence, the original sound of half of the video clips was substituted with the sound of the next measure using *Windows Movie Maker 2.6*.

The 396 stimuli were divided into two groups according to the instrument being played and were presented to 20 musicians attending Conservatory classes (from pre-academic to master level). Judges evaluated whether the sound-gesture video clip combinations were correct using a Likert 3 point scale (2 = congruent; 1 = I am unsure; 0 = incongruent). Judges evaluated only video clips relative to the instrument they knew, i.e., violinists judged only violin video clips and clarinetists judged only clarinet video clips. Aim of the validation test was to ensure that the incongruent clips were easily identifiable by a skilled musician. Videoclips that were incorrectly categorized by more than 5 judges were considered insufficiently reliable and were discarded from the final set of stimuli. A total of 7.5% of the violin stimuli and 6.6% of the clarinet stimuli were discarded. Based on the stimulus validation, 188 congruent (97 clarinet, 91 violin) and 180 incongruent (88 clarinet, 92 violin) videoclips were selected for the EEG study.

The video stimulus size was 15 × 12 cm with a visual angle of 7° 30' 6". Each video was presented for 3000 ms (corresponding to the individual video clip length) against a black background at the center of a high-resolution computer screen. The inter-stimulus interval was 1500 ms. The participants were comfortably seated in a dimly lit test area that was acoustically and electrically shielded. The PC screen was placed 114 cm in front of their eyes. The participants were instructed to gaze at the center of the screen where a small dot served as a fixation point to avoid any eye or body movement during the recording session. All stimuli were presented in random order at the center of the screen in 16 different, randomly mixed, short runs (8 violin video sequences and 8 clarinet video sequences) lasting approximately 3 minutes (plus 2 training sequences). Stimuli presentation and triggering was performed using *Evoke* Software for audiovisual presentation (*ANT Software*, Enschede, The Netherlands). Audio stimulation was administered via a set of headphones.

To keep the subject focused on visual stimulation and ensure the task was feasible for all groups, all participants were instructed and trained to respond as accurately and quickly as possible by pressing a response key with the index or the middle finger corresponding to a 1-note or 2-note stimuli, respectively. The left and right hands were used alternately throughout the recording session, and the order of the hand and task conditions were counterbalanced across participants.

All participants were unaware of the study's aim and of the stimuli properties. At the end of the EEG recording, musicians reported some awareness of their own instrument's audiovisual incongruence, whereas naive individuals showed no awareness of this manipulation.

EEG recordings and analysis. The EEG was recorded and analyzed using *EEProbe* recording software (*ANT Software*, Enschede, The Netherlands). EEG data were continuously recorded from 128 scalp sites according to the 10–5 International System⁶⁴ at a sampling rate of 512 Hz. Horizontal and vertical eye movements were also recorded, and linked ears served as the reference lead. Vertical eye movements were recorded using two electrodes placed below and above the right eye, whereas horizontal movements were recorded using electrodes placed at the outer canthi of the eyes, via a bipolar montage. The EEG and electro-oculogram (EOG) were filtered with a half-amplitude band pass of 0.016–100 Hz. Electrode impedance was maintained below 5 KOhm. EEG epochs were synchronized with the onset of stimulus presentation and analyzed using *ANT-EEProbe software*. Computerized artifact rejection was performed prior to averaging to discard epochs in which eye movements, blinks, excessive muscle potentials or amplifier blocking occurred. The artifact rejection criterion was a peak-to-peak amplitude exceeding 50 mV and resulted in a rejection rate of ~5%. Evoked response potentials (ERPs) from 100 ms before stimulus onset to 3000 ms after stimulus onset were averaged off-line. ERP components were measured when and where they reached their maximum amplitudes. The electrode sites and time windows for measuring and quantifying the ERP components of interest were based on previous literature. The electrode selection

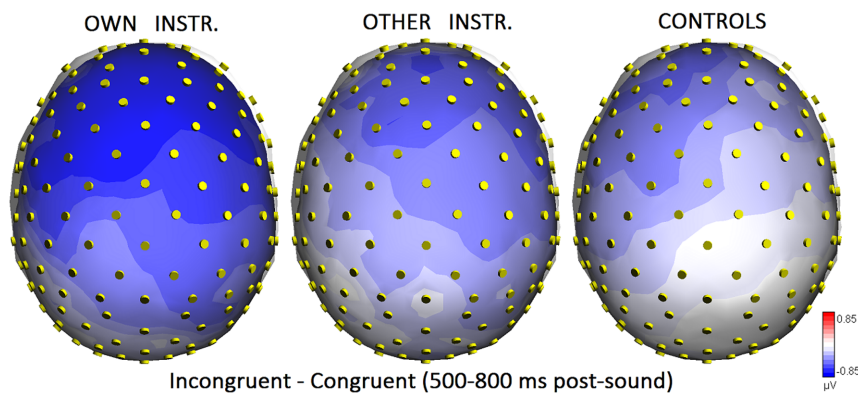


Figure 7 | Top view of isocolor topographic maps computed by plotting the mean voltages of the N400 difference waves for the 3 groups of participants (musicians with their own instrument, musicians with the other instrument, and control subjects).

for the N400 response was also justified by previous studies indicating an anterior scalp distribution for action-related N400 responses^{40,41,65}. The N400 mean area was quantified in the time window corresponding to the maximum amplitude of the differential effect of the mismatch (Incongruent – Congruent). Fig. 7 shows the anterior scalp topography of the difference waves obtained by subtracting ERPs to congruent clips from ERP to incongruent clips in the 3 groups at peak of N400 latency.

It is important to note that each movie clip lasted 3 seconds but during the first second the musician just placed his hands/mouth in correct position to perform the sound. Subsequently, the real sound-gesture onset corresponded 1000 ms after the start of the videoclips.

The peak latency and amplitude of the N170 response were recorded at occipital/temporal sites (PPO9h, PPO10h) between 140–200 ms post-stimulus.

The mean area amplitude of the N400-like response was measured at the fronto-central sites (FCz, Cz, and CPz) in the 1500–2000 ms time window. Multifactorial repeated-measure ANOVAs were applied to the N400 amplitude mean values. The factors of variance were as follows: 1 between-group factor (Groups: musicians and naïve subjects) and 3 within-group factors: Instrument (own instrument or other instrument), Condition (congruent or incongruent), Electrode (depending on the ERP component of interest), and Hemisphere (left hemisphere (LH) or right hemisphere (RH)).

Low-resolution electromagnetic tomography (LORETA) was performed on the ERP waveforms at the N400 latency stage (1500–2000 ms). LORETA is an algorithm that provides discrete linear solutions to inverse EEG problems. The resulting solutions correspond to the 3D distribution of neuronal electrical activity that has the maximally similar orientation and strength between neighboring neuronal populations (represented by adjacent voxels). In this study, an improved version of this algorithm, the standardized weighted (sw) LORETA was used⁶⁶. This version, referred to as swLORETA, incorporates a singular value decomposition-based lead field-weighting method. The source space properties included a grid spacing (the distance between two calculation points) of five points (mm) and an estimated signal-to-noise ratio, which defines the regularization where a higher value indicates less regularization and therefore less blurred results, of three. The use of a value of 3–4 for the computation of the SNR in Tikhonov's regularization produces superior accuracy of the solutions for any inverse problem that is assessed. swLORETA was performed on the grand-averaged group data to identify statistically significant electromagnetic dipoles ($p < 0.05$) in which larger magnitudes correlated with more significant activation. The data were automatically re-referenced to the average reference as part of the LORETA analysis. A realistic boundary element model (BEM) was derived from a T1-weighted 3D MRI dataset through segmentation of the brain tissue. This BEM model consisted of one homogeneous compartment comprising 3446 vertices and 6888 triangles. *Advanced Source Analysis* (ASA) employs a realistic head model of three layers (scalp, skull, and brain) and is created using the BEM. This realistic head model comprises a set of irregularly shaped boundaries and the conductivity values for the compartments between them. Each boundary is approximated by a number of points, which are interconnected by plane triangles. The triangulation leads to a more or less evenly distributed mesh of triangles as a function of the chosen grid value. A smaller value for the grid spacing results in finer meshes and vice versa. With the aforementioned realistic head model of three layers, the segmentation is assumed to include current generators of brain volume, including both gray and white matter. Scalp, skull, and brain region conductivities were assumed to be 0.33, 0.0042, and 0.33, respectively. The source reconstruction solutions were projected onto the 3D MRI of the Collins brain, which was provided by the Montreal Neurological Institute. The probabilities of source activation based on Fisher's *F*-test were provided for each independent EEG source, whose values are indicated in a "unit" scale (the larger the value, the more significant). Both the segmentation and generation of the head model were performed using the ASA software program *Advanced Neuro Technology* (ANT, Enschede, Netherlands)⁶⁷.

Response times exceeding the mean ± 2 standard deviations were excluded. Hit and miss percentages were also collected and arc sin transformed to allow for sta-

tistical analyses. Behavioral data (both response speed and accuracy data) were subjected to multifactorial repeated-measures ANOVA with factors for group (musicians, $N = 12$; controls, $N = 12$) and condition (congruence, incongruence). A Tukey's test was used for post-hoc comparisons among means.

- Kohler, E. *et al.* Hearing sounds, understanding actions: action representation in mirror neurons. *Science* **297**, 846–848 (2002).
- Keysers, C. *et al.* Audiovisual mirror neurons and action recognition. *Exp Brain Res.* **153**, 628–36 (2003).
- Bangert, M. & Altenmüller, E. O. Mapping perception to action in piano practice: a longitudinal DC-EEG study. *BMC Neurosci.* **15**, 4:26 (2003).
- Janata, P. & Grafton, S. T. Swinging in the brain: shared neural substrates for behaviors related to sequencing and music. *Nat Neurosci.* **6**, 682–7 (2003).
- Lahav, A., Saltzman, E. & Schlaug, G. Action representation of sound: audiomotor recognition network while listening to newly acquired actions. *J Neurosci.* **27**, 308–14 (2007).
- McGurk, H. & MacDonald, J. Hearing lips and seeing voices. *Nature* **264**, 746–748 (1976).
- Calvert, G. A., Campbell, R. & Brammer, M. J. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol.* **10**, 649–657 (2000).
- Jones, J. A. & Callan, D. E. Brain activity during audiovisual speech perception: An fMRI study of the McGurk effect. *NeuroReport* **14**, 1129–1133 (2003).
- Näätänen, R. The mismatch negativity - a powerful tool for cognitive neuroscience. *Ear & Hearing* **16**, 6–18 (1995).
- Sams, M. *et al.* Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex. *Neurosci. Lett.* **127**, 141–145 (1991).
- Colin, C., Radeau, M., Soquet, A. & Deltenre, P. Generalization of the generation of an MMN by illusory McGurk percepts: Voiceless consonants. *Clin. Neuropsychol.* **115**, 1989–2000 (2004).
- Kislyuk, D., Mottonen, R. & Sams, M. Visual Processing Affects the Neural Basis of Auditory Discrimination. *J Cogn Neurosci.* **20**, 2175–2184 (2008).
- Besle, J. *et al.* Visual Activation and Audiovisual Interaction in the Auditory Cortex during Speech Perception: Intracranial Recordings in Humans. *J. Neurosci.* **28**, 14301–14310 (2008).
- Mottonen, R., Jarvelainen, J., Sams, M. & Hari, R. Viewing speech modulates activity in the left SI mouth cortex. *NeuroImage.* **24**, 731–737 (2004).
- Pantev, C., Lappe, C., Herholz, S. C. & Trainor, L. Auditory-somatosensory integration and cortical plasticity in musical training. *Ann. N.Y. Acad. Sci.* **1169**, 143–50 (2008).
- Pantev, C. *et al.* Music and learning-induced cortical plasticity. *Ann. N.Y. Acad. Sci.* **99**, 438–50 (2003).
- Hasegawa, T. *et al.* Learned audio-visual cross-modal associations in observed piano playing activate the left planum temporale. An fMRI study. *Brain Res. Cogn. Brain Res.* **20**, 510–518 (2004).
- Schuboz, R. I. & Von Cramon, D. Y. A blueprint for target motion: fMRI reveals perceived sequential complexity to modulate premotor cortex. *NeuroImage* **16**, 920–935 (2002).
- McIntosh, A., Cabeza, R. E., Lobaugh, N. J. Analysis of neural interactions explains the activation of occipital cortex by an auditory stimulus. *J. Neurophysiol.* **80**, 2790–2796 (1998).
- Barrett, K. C., Ashley, R., Strait, D. L. & Kraus, N. Art and Science: How Musical Training Shapes the Brain. *Fron. Psychol.* **4**, 713 (2013).
- Schlaug, G. The brain of musicians. A model for functional and structural adaptation. *Ann. N.Y. Acad. Sci.* **930**, 281–299 (2001).
- Gaser, C. & Schlaug, G. Gray matter differences between musicians and non-musicians. *Ann. N.Y. Acad. Sci.* **999**, 514–517 (2003).
- Gaser, C. & Schlaug, G. Brain structures differ between musicians and non-musicians. *J. Neurosci.* **23**, 9240–9245 (2003).



24. Lee, D. J., Chen, Y. & Schlaug, G. Corpus callosum: musician and gender effects. *NeuroReport* **14**, 205–209 (2003).
25. Bermudez, P. & Zatorre, R. J. Differences in gray matter between musicians and non-musicians. *Ann. N.Y. Acad. Sci.* **1060**, 395–399 (2005).
26. Pantev, C. *et al.* Increased auditory cortical representation in musicians. *Nature* **392**, 811–814 (1998).
27. Proverbio, A. M., Manfredi, M., Zani, A. & Adorni, R. Musical expertise affects neural bases of letter recognition. *Neuropsychologia* **51**, 538–49 (2013).
28. Kraus, N. & Chandrasekaran, B. Music training for the development of auditory skills. *Nat. Rev. Neurosci.* **11**, 599–605 (2010).
29. Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B. & Taub, E. Increased cortical representation of the fingers of the left hand in string players. *Science* **270**, 305–307 (1995).
30. Hund-Georgiadis, M. & Von Cramon, D. Y. Motor learning-related changes in piano players and nonmusicians revealed by functional magnetic-resonance signal. *Exp. Brain Res.* **125**, 417–425 (1999).
31. Bangert, M. & Schlaug, G. Specialization of the specialized in features of external human brain morphology. *Eur. J. Neurosci.* **24**, 1832–1834 (2006).
32. Zatorre, R. J., Chen, J. L. & Penhune, V. B. When the brain plays music: auditory-motor interactions in music perception and production. *Nat. Rev. Neurosci.* **8**, 547–558 (2007).
33. Schulz, M., Ross, B. & Pantev, C. Evidence for training-induced crossmodal reorganization of the cortical functions in trumpet players. *Neuroreport* **14**, 157–161 (2003).
34. Baumann, S. *et al.* A network for audio-motor coordination in skilled pianists and non-musicians. *Brain Res.* **1161**, 65–78 (2007).
35. Schmithorst, V. J. & Wilke, M. Differences in white matter architecture between musicians and non-musicians: a diffusion tensor imaging study. *Neurosci Lett.* **15**, 57–60 (2002).
36. Bengtsson, S. L. *et al.* Extensive piano practicing has regionally specific effects on white matter development. *Nat Neurosci.* **8**, 1148–50 (2005).
37. Winkler, I. & Czigler, I. Evidence from auditory and visual event-related potential (ERP) studies of deviance detection (MMN and vMMN) linking predictive coding theories and perceptual object representations. *Int J Psychophysiol.* **83**, 132–43 (2012).
38. Paraskevopoulos, E., Kuchenbuch, A., Herholz, S. C. & Pantev, C. Musical expertise induces audiovisual integration of abstract congruency rules. *J Neurosci.* **32**, 18196–203 (2012).
39. Gunter, T. C. & Bach, P. Communicating hands: ERPs elicited by meaningful symbolic hand postures. *Neurosci Lett.* **30**, 52–6 (2004).
40. Proverbio, A. M., Crotti, N., Manfredi, M., Adorni, R. & Zani, A. Who needs a referee? How incorrect basketball actions are automatically detected by basketball players' brain. *Sci. Rep.* **2**, 883 (2012).
41. Proverbio, A. M. & Riva, F. RP and N400 ERP components reflect semantic violations in visual processing of human actions. *Neurosci. Lett.* **459**, 142 (2009).
42. Sitnikova, T., Holcomb, P. J., Kiyonaga, K. A. & Kuperberg, G. R. Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. *J Cogn. Neurosci.* **20**, 11–22 (2008).
43. Proverbio, A. M., Calbi, M., Zani, A. & Manfredi, M. Comprehending body language and mimics: an ERP and neuroimaging study on Italian actors and viewers. *PlosOne* **7**, e91294 (2014).
44. van Elk, M., van Schie, H. T. & Bekkering, H. Semantics in action: an electrophysiological study on the use of semantic knowledge for action. *J Physiol Paris.* **102**, 95–100 (2008).
45. Amoroso, L. *et al.* N400 ERPs for actions: building meaning in context. *Front Hum Neurosci.* **4**, 7:57 (2013).
46. Näätänen, R. The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behav. Brain Sci.* **13**, 201–288 (1990).
47. Näätänen, R., Brattico, E. & Tervaniemi, M. [Mismatch negativity (MMN): A probe to auditory cognition and perception in basic and clinical research] *The cognitive electrophysiology of mind and brain* [Zani, A., Proverbio, A. M. (eds.)] [343–355], (Academic Press, San Diego 2003).
48. Hamilton, A. F. & Grafton, S. T. Action Outcomes Are Represented in Human Inferior Frontoparietal Cortex. *Cereb. Cortex* **18**, 160–1168 (2008).
49. Rizzolatti, G. *et al.* Localization of grasp representations in humans by PET: 1. Observation versus execution. *Exp Brain Res.* **111**, 246–52 (1996).
50. Urgesi, C., Berlucchi, G. & Aglioti, S. M. Magnetic Stimulation of Extrastriate Body Area Impairs Visual Processing of Nonfacial Body Parts. *Curr. Biol.* **14**, 2130–2134 (2004).
51. Watkins, K. E., Straffella, A. P. & Paus, T. Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia.* **41**, 989–994 (2003).
52. Fadiga, L., Craighero, L., Buccino, G. & Rizzolatti, G. Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *Eur. J. Neurosci.* **15**, 399–402 (2002).
53. Nishitani, N. & Hari, R. Viewing lip forms: cortical dynamics. *Neuron* **36**, 1211–1220 (2002).
54. D'Ausilio, A., Altenmüller, E., Olivetti Belardinelli, M. & Lotze, M. Cross-modal plasticity of the motor cortex while listening to a rehearsed musical piece. *Eur J Neurosci.* **24**, 955–958 (2006).
55. Bangert, M. *et al.* Shared networks for auditory and motor processing in professional pianists: evidence from fMRI conjunction. *NeuroImage* **30**, 917–926 (2006).
56. Calvo-Merino, B., Glaser, D. E., Grezes, J., Passingham, R. E. & Haggard, P. Action observation and acquired motor skills: an fMRI study with expert dancers. *Cereb Cortex* **15**, 1243–1249 (2005).
57. Grill Spector, K., Knouf, N. & Kanwisher N. The fusiform face area subserves face perception, not generic within-category identification. *Nat Neurosci.* **7**, 555–62 (2004).
58. Schwarzklose, R. F., Baker, C. I. & Kanwisher, N. Separate face and body selectivity on the fusiform gyrus. *J Neurosci.* **23**, 11055–9 (2005).
59. Giard, M. H., Perrin, F., Pernier, J. & Bouchet, P. Brain generators implicated in processing of auditory stimulus deviance: A topographic event-related potential study. *Psychophysiol.* **27**, 627–640 (1990).
60. Rinne, T., Ilmoniemi, R. J., Sinkkonen, J., Virtanen, J. & Näätänen, R. Separate time behaviors of the temporal and frontal MMN sources. *Neuroimage.* **12**, 14–19 (2000).
61. Schröger, E. A neural mechanism for involuntary attention shifts to changes in auditory stimulation. *J Cogn Neurosci.* **8**, 527–539 (1996).
62. Alho, K., Escera, C., Diaz, R., Yago, E. & Serrà, J. M. Effects of involuntary auditory attention on visual task performance and brain activity. *NeuroReport.* **8**, 3233–3237 (1997).
63. Zani, A. & Proverbio, A. M. *The cognitive Electrophysiology of mind and brain* [Zani, A., Proverbio, A. M. (eds.)] (Academic Press, San Diego 2003).
64. Oostenveld, R. & Praamstra, P. The five percent electrode system for high-resolution EEG and ERP measurements. *Clin Neurophysiol.* **112**, 713–719 (2001).
65. Proverbio, A. M., Calbi, M., Manfredi, M. & Zani, A. Comprehending body language and mimics: an ERP and neuroimaging study on Italian actors and viewers. *Plos ONE.* **9**, e91294 (2014).
66. Palmero-Soler, E., Dolan, K., Hadamschek, V. & Tass, P. A. SwLORETA: A novel approach to robust source localization and synchronization tomography. *Phys. Med. Biol.* **52**, 1783–1800 (2007).
67. Zanow, F. & Knosche, T. R. ASA-Advanced Source Analysis of continuous and event-related EEG/MEG signals. *Brain Topog.* **16**, 287–290 (2004).

Acknowledgments

We are very grateful to all participants for their help, particularly the students of *Conservatorio Giuseppe Verdi* of Milan and their teachers. Special thanks go to M^o Mauro Loguercio for writing the music played in this study. We also wish to thank Andrea Orlandi, Francesca Pisanu and Giancarlo Caramenti of the Institute for Biomedical Technologies of CNR for their assistance in this study.

Author contributions

A.M.P. designed the methods and experiments. M.M. and M.C. performed data acquisition and analyses. A.M.P. interpreted the results and wrote the paper. A.Z. co-worked on source localization analysis and interpretation. All authors have contributed to, seen and approved the manuscript.

Additional information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Proverbio, A.M., Calbi, M., Manfredi, M. & Zani, A. Audio-visuomotor processing in the Musician's brain: an ERP study on professional violinists and clarinetists. *Sci. Rep.* **4**, 5866; DOI:10.1038/srep05866 (2014).



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>