SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
NEUROIMAGING
NEUROSCIENCE
SENSORY SYSTEMS
BEHAVIOUR

Received
5 April 2011

Accepted
20 July 2011

Published
4 August 2011

Correspondence and requests for materials should be addressed to A.M.P. (mado.proverbio@unimib.it)

When a photograph can be heard: Vision activates the auditory cortex within 110 ms

Alice Mado Proverbio[1], Guido Edoardo D'Aniello[1], Roberta Adorni[1] & Alberto Zani[2]

[1]Dept. of Psychology, University of Milano-Bicocca, Piazza dell'Ateneo Nuovo 1, 20126 Milan, Italy, [2]Institute of Molecular Bioimaging and Physiology, National Research Council (CNR), Milano-Segrate, Italy.

As the makers of silent movies knew well, it is not necessary to provide an actual auditory stimulus to activate the sensation of sounds typically associated with what we are viewing. Thus, you could almost hear the neigh of Rodolfo Valentino's horse, even though the film was mute. Evidence is provided that the mere sight of a photograph associated with a sound can activate the associative auditory cortex. High-density ERPs were recorded in 15 participants while they viewed hundreds of perceptually matched images that were associated (or not) with a given sound. Sound stimuli were discriminated from non-sound stimuli as early as 110 ms. SwLORETA reconstructions showed common activation of ventral stream areas for both types of stimuli and of the associative temporal cortex, at the earliest stage, only for sound stimuli. The primary auditory cortex (BA41) was also activated by sound images after ~ 200 ms.

Neuroimaging data[1–3] have shown the existence of audiomotor multisensory neurons in the posterior region of the superior temporal sulcus (pSTS) and in the middle temporal gyrus (MTG) that respond to the sounds and visual images of objects and animals; these regions also respond to letters and speech sounds and labial movements[4]. In addition, these regions are activated more strongly by audiovisual stimuli than by unisensory stimuli, thus suggesting multisensory integration of inputs from two modalities[5]. This multisensory integration is particularly strong for linguistic stimuli, in that an incongruent visual stimulus can qualitatively change the auditory perception at the level of the auditory cortex[6–8]. In monkeys, audiovisual "mirror" neurons have been discovered in the ventral premotor cortex[9,10]. These neurons discharge both when the animal performs a specific action and when it either hears the sound associated with that action or sees the action.

With regard to the timing of this integration, in an electrophysiological study by Senkowski[11], processing of multisensory (audiovisual) and unisensory (auditory or visual) stimuli were explored using naturalistic water splash sounds and corresponding visual images. They found an early effect of multisensory integration (120–140 ms) over the posterior brain areas; this was followed by later (210–350 ms) activity involving (among other areas) the temporal cortex (MTG and STG).

With the exception of direct neurophysiological evidence of "*audiovisual mirror neurons*" (in monkeys), most, if not all, neuroimaging studies of multisensory interactions in humans have relied on estimating audiovisual interactions by comparing the response to the multisensory stimulus and a combination of the responses to the unisensory stimuli presented in isolation. In the present study, the subjects received no auditory stimulation, but rather received only visual stimuli consisting of scenes strongly linked (or not linked) to a sound association (as estimated by an independent group of viewers); these included an image of a man playing a trumpet or an image of a sleeping child. All of the images (see Fig. 6 for some examples) were carefully matched for their size, average luminance, luminance profile, affective value and presence of animals or humans and differed only in their degree of auditory content. High-density EEG was recorded from 15 right-handed volunteers, and swLORETA was performed on the brain activity related to sound and non-sound processing, as well as on their differential activation.

Results

Occipital P1 was not affected by stimulus category, neither in latency ($F1,13 = 0.009$; $p = 0.93$; sound = 105 ms, non-sound = 105 ms), nor in amplitude ($F1,13 = 0.0003$; $p = 0.99$; sound = 6.52 μV, non-sound = 6.52 μV), as clearly appreciable by looking at ERP waveforms of Fig. 1 (Top) and relative topographical maps (Bottom).

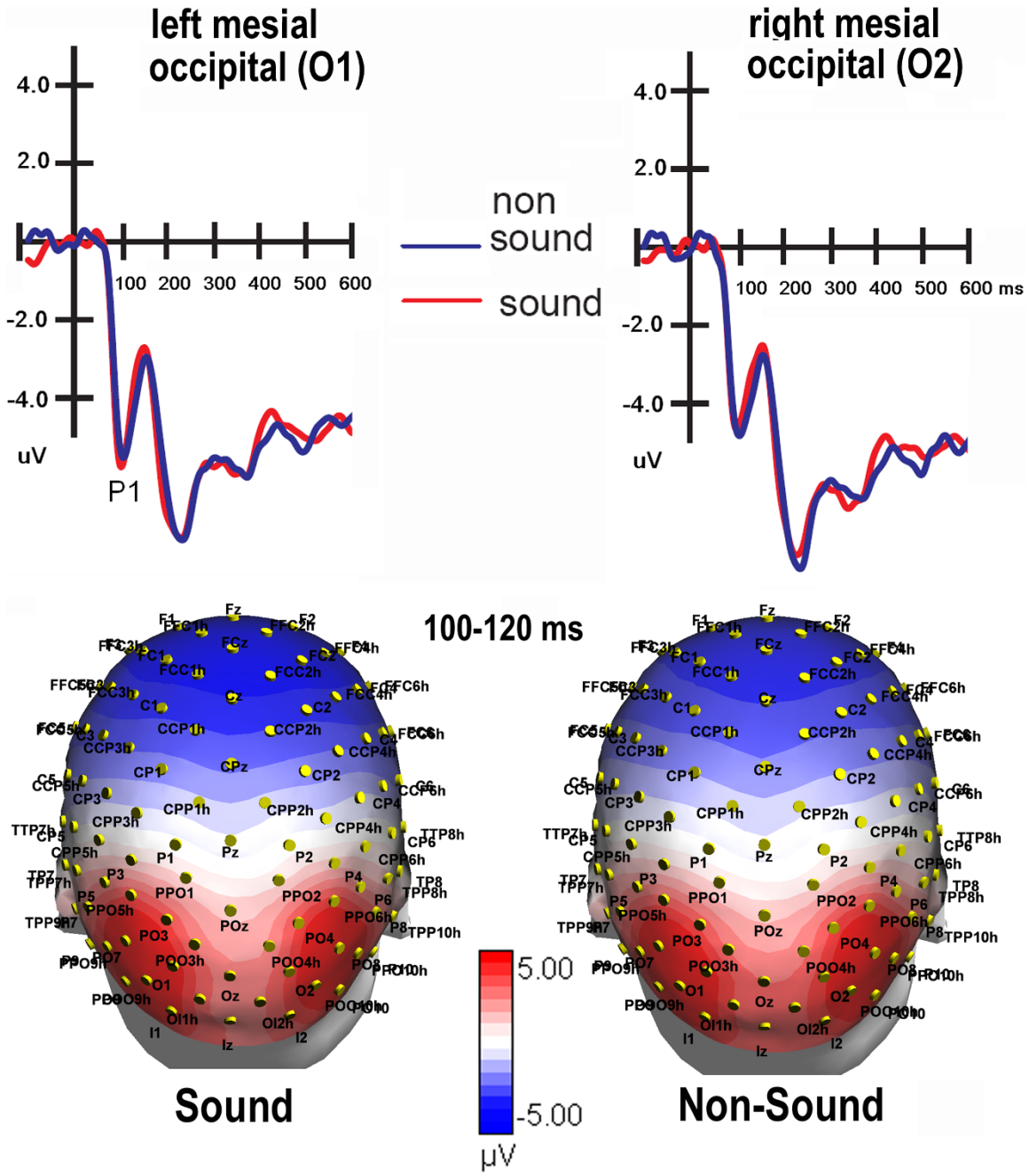

Figure 1 | (Top) Grand-average ERP waveforms recorded at left and right mesial occipital sites in response to sound and non-sound stimuli. (Bottom) Topographical maps obtained by plotting the colour-coded average voltage recorded in the 100–120 time window in response to sound and non-sound stimuli. It can be appreciated that, while both waveforms and maps relative to the early sensory visual activity (P1) were not affected whatsoever by stimulus content, sound stimuli elicited a stronger negativity (N1) having a fronto-central distribution.

Frontal N1 was differentially affected by stimulus category ($F_{1,13} = 4.44$; $p < 0.05$; $\epsilon = 1$), being larger in response to sound stimuli than to non-sound stimuli (sound = −3.41 μV, SE = 0.51; non-sound = −2.74 μV, SE = 0.38); this is illustrated in the waveforms shown in Fig. 2. N1 reached its maximum amplitude at central (C3, C4) sites ($F(2, 20) = 18.5$; $p < 0.00005$; $\epsilon = 0.31$). The frontal N2 response was also differentially affected by stimulus content ($F_{1,13} = 4.87$; $p < 0.045$; $\epsilon = 1$), having a greater amplitude in response to sound stimuli than to non-sound stimuli (sound = −4.94; μV, SE = 1.08; non-sound = −4.36 μV; SE = 1.02), as showed in topographical maps of Fig. 1 (Bottom). N2 reached its maximum amplitude at central (C3, C4) sites ($F2, 25 = 19.7$; $p<0.00001$; $\epsilon = 0.39$). To identify the intracranial sources of the increased bioelectrical activity elicited by sound stimuli, two swLORETAs (displayed in Fig. 3) were applied to the difference voltages obtained by subtracting ERPs to non-sound from ERPs to sound stimuli in the two time windows of 100–120 ms (corresponding to N1 peak), and 205–225 ms (corresponding to N2 peak). The results are reported in Table 1, showing a list of electromagnetic dipoles explaining the difference voltages, along with their Talairach coordinates. In the first time window it was found an activation of the left MTG (BA21), along with the right MOG and medial frontal gyrus. After about 100 ms the signal power was stronger, and included the activation of the left middle frontal gyrus, the right STG (BA38), the left ITG (BA20), and the left STG (BA41), the latter corresponding to the primary auditory cortex.

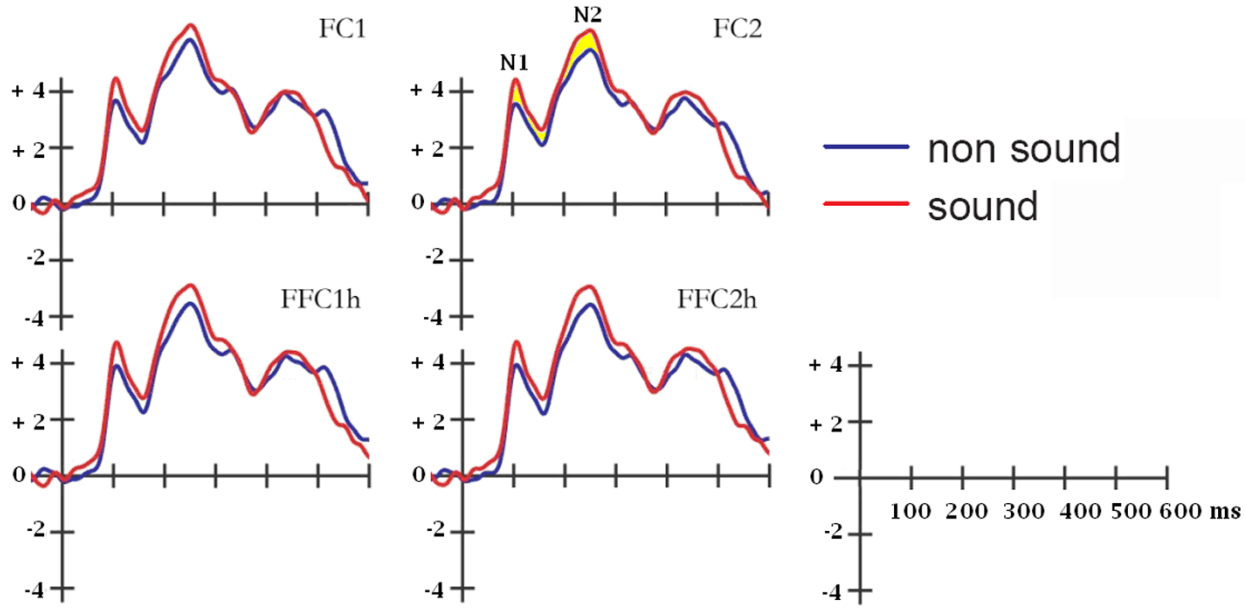

Figure 2 | **Grand-average ERP waveforms recorded at left and right fronto-central sites in response to sound and non-sound stimuli.**

The later P3 response (600–800 ms) was larger in response to sound stimuli than to non-sound stimuli ($F1,13 = 5.97$; $p < 0.042$). The significant interaction of stimulus category x hemisphere ($F1,13 = 5.1$; $p < 0.03$) and relative post-hoc comparisons were indicative of larger sound vs. non-sound differences over the left hemisphere (LH) compared with the right hemisphere (RH: sound = 1.66, non-sound = 1.39 μV; LH: sound = 1.86, non-sound = 1.25 μV), as shown in Fig. 4.

To locate the possible neural source of the auditory content effect, two different swLORETA source reconstructions were performed independently for the sound and non-sound stimuli during the 600–800-ms time window, which corresponds to the peak of the temporal P3. The inverse solution is displayed in Fig. 5 and shows that the processing of both stimuli classes was associated with a common set of left and right generators (listed in Table 2) located in the ventral stream and devoted to both object/face processing (e.g., BA20 and BA37) and scene encoding. However, only perceived sound stimuli activated the superior temporal gyrus (BA38). In order to ascertain which regions were more robustly activated specifically during sound processing in the P3 latency range, an additional

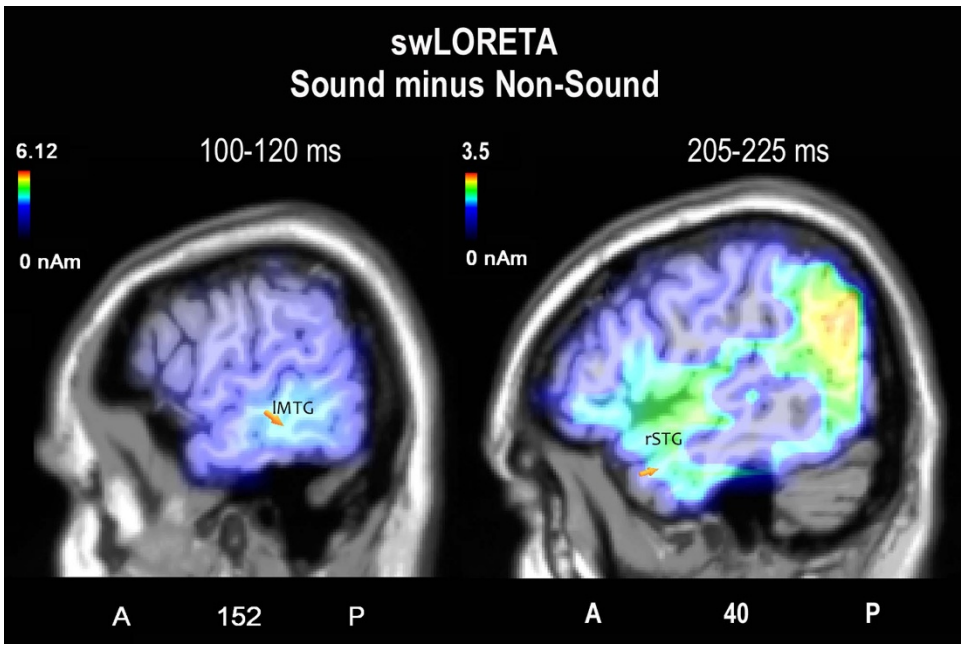

Figure 3 | **Sagittal view of intra-cranial active sources explaining the difference voltage sound – non-sound stimuli computed for the two time windows of 100–120 ms (corresponding to N1 peak) and 205–225 ms (corresponding to N2 peak).** The different colours represent differences in the magnitude of the electromagnetic signal (in nAm). The electromagnetic dipoles are shown as arrows and indicate the position, orientation and magnitude of dipole modelling solution applied to the ERP waveform in the specific time window. The two sagittal sections are centred on the left MTG (BA21) and the right STG (BA38), respectively. L = left; R = right; numbers refer to the displayed brain slice in sagittal view. The first is a left hemispheric view, the second is a right hemispheric view.

Table 1 | Talairach coordinates corresponding to the intracortical generators, which explain the surface voltage recorded during the 100–120 and 205–225 ms time windows, respectively, in response to sound and non-sound stimuli. Magnitude is expressed in nAm; H = hemisphere; BA = Brodmann area.

SOUND MINUS NON-SOUND DFFERENCE						
Magnit	T-x [mm]	T-y [mm]	T-z [mm]	H	Gyrus	BA
100–120 ms						
6.40	31	−88.3	3	R	Middle Occipital Gyrus	18
3.68	**−58.5**	**−25.5**	**−8.1**	**L**	**Middle Temporal Gyrus**	**21**
3.67	1.5	38.2	−17.9	R	Medial Frontal Gyrus	11
205–225 ms						
5.43	31	−89.3	11.9	R	Middle Occipital Gyrus	19
4.27	31	−73	49.2	R	Superior Parietal Lobule	7
4.10	−8.5	−98.5	2.1	L	Cuneus	
3.36	−8.5	1.4	38.3	L	Cingulate Gyrus	24
2.76	**50.8**	**8.2**	**−20**	**R**	**Superior Temporal Gyrus**	**38**
2,16	**−38,5**	**−28,5**	**17,1**	**L**	**Superior Temporal Gyrus**	**41**
3.43	−28.5	21.4	40	L	Middle Frontal Gyrus	8
2.38	1.5	38.2	−17.9	R	Medial Frontal Gyrus	11
2.37	−48.5	−0.6	−28.2	L	Inferior Temporal Gyrus	20
2.23	40.9	2.4	29.4	R	Precentral Gyrus	6

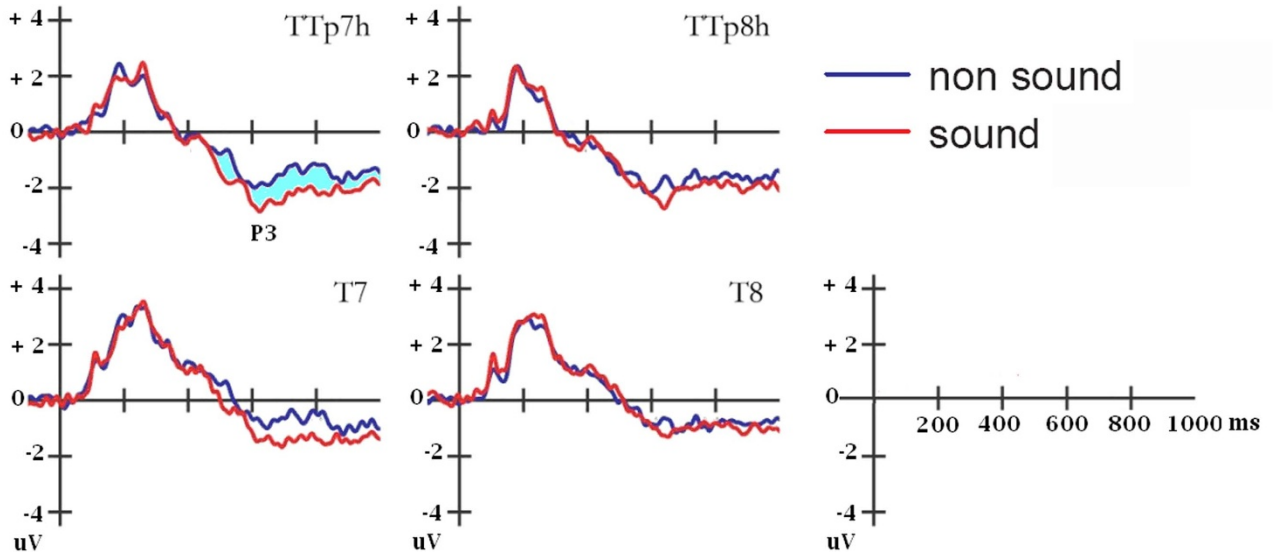

Figure 4 | Grand-average ERP waveforms recorded at left and right temporo-parietal and posterior-temporal sites in response to sound and non-sound stimuli.

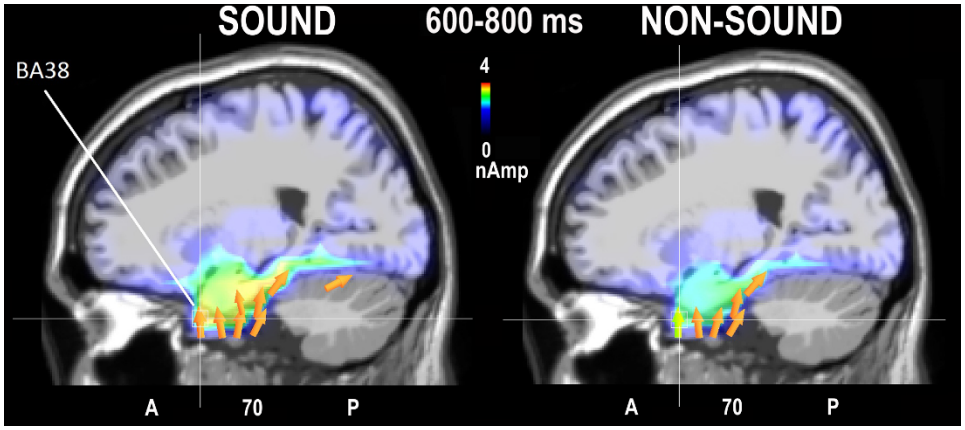

Figure 5 | Sagittal view of intra-cranial active sources for the processing of sound (left) and non-sound stimuli (right) according to the swLORETA analysis during the 600–800-ms time window. Evident is a stronger sound-related temporal activation, which likely reflects the processing of sound objects.

Table 2 | Talairach coordinates corresponding to the intracortical generators, which explain the surface voltage recorded during the 600–800-ms time window in response to sound and non-sound stimuli. Magnitude is expressed in nAmp; H = hemisphere; BA = Brodmann area.

Magnit	T-x [mm]	T-y [mm]	T-z [mm]	H	Gyrus	BA
SOUND						
14.08	50.8	−33.7	−23.6	R	Fusiform Gyrus	20
13.95	−18.5	−8.0	−28.9	L	Uncus	36
13.93	50.8	−0.6	−28.2	R	Middle Temporal Gyrus	21
13.87	50.8	−16.1	−22.2	R	Fusiform Gyrus	20
13.85	21.2	−24.5	−15.5	R	Parahipp.Gyrus	35
13.84	31.0	9.1	−27.5	R	Superior Temporal Gyrus	38
13.85	40.9	−55.0	−17.6	R	Fusiform Gyrus	37
13.81	−48.5	−55.0	−17.6	L	Fusiform Gyrus	37
13.72	31.0	−15.3	−29.6	R	Uncus	20
13.19	40.9	−75.2	−19.1	R	Posterior Lobe,	
13.14	−58.5	−8.7	−21.5	L	Inferior Tempoal Gyrus	20
13.04	−38.5	−88.3	3.0	L	Middle Occipital Gyrus	18
NON-SOUND						
13.441	50.8	−33.7	−23.6	R	Fusiform Gyrus	20
13.354	50.8	−55.0	−17.6	R	Fusiform Gyrus	37
13.260	−48.5	−55.0	−17.6	L	Fusiform Gyrus	37
13.250	−18.5	−8.0	−28.9	L	Uncus	36
13.217	50.8	−0.6	−28.2	R	Middle Temporal Gyrus	21
13.210	21.2	−24.5	−15.5	R	Parahipp. Gyrus	35
13.171	31.0	−15.3	−29.6	R	Uncus	20
13.164	50.8	−16.1	−22.2	R	Fusiform Gyrus	20
13.112	31.0	9.1	−27.5	R	Superior Temporal Gyrus	38

swLORETA was computed for the difference signals obtained by subtracting the bio-electric non-sound activity from the sound activity recorded during the 600–800-ms time window. The electromagnetic dipoles (listed in Table 3) represent intra-cranial sources of activity that were significantly stronger in response to sound than non-sound stimuli; the ITG, MTG and STG cortices (BA20, 21 and 38, respectively) were among the strongest foci.

Discussion

This early effect of multisensory integration is consistent with previous reports comparing multisensory audiovisual stimuli with unimodal visual or auditory stimuli[11, 12].

The lack of any visual sensory stimulus-dependent modulation of ERPs suggests than the differences found between sound vs. non-sound stimuli were not due to their perceptual characteristics, but, very likely, to the auditory content of visual information carried out by sound stimuli.

As for the earliest effect at N1 level (100–120 ms), the inverse solution applied to the difference voltage sound minus non-sound showed that the main sources of activity for this effect were not entirely visual (MTG, MOG, rMFG). It cannot be excluded that the early right medial frontal activation reflected an attention modulation, besides multisensory integration processes. However the role of medial frontal cortex in auditory processing has also been established. For example, Anderer et al.[13] applied LORETA source reconstruction to auditory ERPs recorded in an oddball task, finding an activation of the superior temporal gyrus [auditory cortex, Brodmann areas (BA) 41, 42, 22] for both N1 and N2 responses and also a medial frontal source (BA 9, 10, 32) for N2 response. An early activation of both occipital, temporal and frontal cortices for multisensory audio-visual (AV) processing was reported by a recent fMRI study[14] in which subjects passively perceived sounds and images of objects presented either alone or simultaneously. After AV stimulation, a significant activity (after 6–7- sec) was observed in superior temporal gyrus, middle temporal gyrus, right occipital cortex, and inferior frontal cortex, besides the right Heschl's gyrus, thus suggesting the crucial role of these areas in object-dependent audio-visual integration.

Table 3 | Intracranial generators relative to the difference signal obtained by subtracting the bio-electric non-sound response from the sound response recorded during the 600–800-ms time window. The listed electromagnetic dipoles represent sources of activity that respond significantly more strongly in response to sound than non-sound stimuli. The strongest responding foci included the right ITG, MTG and STG (BA20, 21 and 38, respectively).

SOUND MINUS NON-SOUND DFFERENCE						
Magnit	T-x [mm]	T-y [mm]	T-z [mm]	H	Gyrus	BA
6.201	50.8	−44.8	−16.9	R	Fusiform Gyrus	37
6.185	**60.6**	**−24.5**	**−15.5**	**R**	**Inferior Temporal Gyrus**	**20**
6.181	**50.8**	**−0.6**	**−28.2**	**R**	**Middle Temporal Gyrus**	**21**
5.991	**31.0**	**9.1**	**−27.5**	**R**	**Superior Temporal Gyrus**	**38**
5.665	−8.5	−0.6	−28.2	L	Uncus	28
4.796	−28.5	−97.5	−5.7	L	Lingual Gyrus	18
4.445	−58.5	−8.7	−21.5	L	Inferior Temporal Gyrus	20
4.436	−58.5	−55.0	−17.6	L	Fusiform Gyrus	37
4.014	21.2	−98.5	2.1	R	Cuneus	18
2.706	1.5	−29.4	26.0	R	Cingulate Gyrus	23

According to Näätänen and Winkler[15], the fronto-central N1 (100 ms) response reflects the initial access to mental auditory representation, whereas the fronto-central N250 (200–250 ms) response indexes the stage of multisensory integration, with visual inputs coming from the ventral stream. Other electrophysiological studies (e.g., Ref. 16) found an increase in anterior N2 amplitude while imaging an auditory stimulus, which likely suggests activation of an auditory mental representation.

Considering the visual and implicit nature of our experiment—the participants were actively looking for target scenes (cycle races) while ignoring other images—our ERP data indicate an automatic and early access to object sound properties. Studies of multimodal integration[11] have suggested an early activation of audiomotor neurons at about 100 ms that is followed by more robust activity in a later time window (210–350 ms). This activity would involve regions of the associative temporal cortex (MTG and STG, among others), as shown by the swLORETA inverse solutions performed on our N1, N2 and P3 data. Interestingly, direct neurophysiological data[9] suggest that the STS is an integration area for visual and auditory inputs (such as the sight of an action and its corresponding sound), thus demonstrating the existence of audiovisual mirror neurons.

In conclusion, we provide evidence that the mere sight of scenes and objects typically associated with sound will automatically activate auditory representation in several regions within the associative temporal and even auditory primary cortex. Moreover, these regions are known to be engaged in the perception of complex sounds[17], audiovisual processing of speech stimuli[18], audiovisual integration[19] and auditory verbal hallucinations[20, 21], which tend to be selectively associated with right STS activation.

Methods

Subjects. Fifteen healthy right-handed university students (8 men and 7 women) participated in this study as unpaid volunteers. They earned academic credit for their participation. Their mean age was 22.8 years, ranging from 20 to 27 years. All had normal or corrected-to-normal vision and reported no history of neurological illness or drug abuse. Their right-handedness and right ocular dominance were confirmed using the Italian version of the Edinburgh Handedness Inventory, a laterality preference questionnaire. All experiments were conducted with the understanding and written consent of each participant. No participant was excluded for technical reasons. The experimental protocol was approved by the ethics committee of the University of Milano-Bicocca.

Stimuli and materials. The stimulus set consisted of 300 complex ecological scenes. The pictures were downloaded from Google Images (the examples reported in Fig 6 are custom-made and copy-right free). The two classes of stimuli (sound and non-sound) were matched for their size (350 × 350 pixels), luminance (41.92 cd/cm^2), affective value and presence of animals or persons. Half of the images (150) evoked a strong auditory image (sound stimuli), whereas the other half were not linked to any particular sound (non-sound stimuli). The stimulus set was selected from a larger set

Figure 6 | Example images of stimuli in the sound and non-sound categories.

of images by presenting them to a group of 20 judges (10 men and 10 women) and asking them to score whether they evoked an auditory association using a 3-point scale (with 2, 1 and 0 being strong, weak and absent auditory content, respectively).

To provide a clear distinction between the sound and non-sound stimulus groups, pictures scoring an average value of 0.5–2 were placed in the sound category, whereas pictures scoring a value of 0 were placed in the non-sound category. A *t*-test applied to the 2 groups confirmed that their auditory contents were significantly different (Sound = 1.41, SE = 0.37; Non-sound = 0; *t*-value = 46.58; $p < 0.05$). Three hundred (150 sound and 150 non-sound) images meeting the above criteria were then selected to create the final stimulus set; some example images are shown in Fig. 6.

The stimuli in the 2 classes were also matched for their affective value by presenting the pictures to a group of 10 judges (5 men and 5 women) different than those used above and asking them to evaluate the stimuli in terms of their affective content using a 3-point scale (with 2, 1 and 0 being strong, weak and null affective value, respectively). A *t*-test applied to the 2 groups confirmed that their affective values were not significantly different (Sound = 0.76; Non-sound = 0.66; *t*-value = 1.68; $p = 0.09$).

Twenty-five additional photos depicting a cycle race were included in the stimulus set for the subjects to perform a secondary task (described below); these images were of similar average luminance, size and spatial distribution as the other images. The sound and non-sound images were presented in random order together with the 25 cycle race photos. The stimulus size was 14.2 × 14.2 cm subtending a visual angle of 6°43′01″. Each image was presented for 1000 ms against a dark grey background at the center of a computer screen with an ISI of 1500–1900 ms.

Task and procedure. The participants were comfortably seated in a darkened test area that was acoustically and electrically shielded. A high-resolution VGA computer screen was placed 120 cm in front of their eyes. The subjects were instructed to gaze at the center of the screen (where a small circle served as a fixation point) and to avoid any eye or body movement during the recording session. The stimuli were presented in random order at the center of the screen in 6 different randomly mixed short runs lasting approximately 2 minutes and 40 seconds. To keep the subject focused on the visual stimuli, the task consisted of responding as accurately and quickly as possible to photos displaying cycle races by pressing a response key with the index finger of the left or right hand; all other photos were to be ignored. The left and right hands were used alternately throughout the recording session, and the order of the hand and task conditions were counterbalanced across the subjects. For each experimental run, the target stimuli varied between 3–7, and the presentation order differed among the subjects.

EEG recording and analysis. The EEG data were continuously recorded from 128 scalp sites at a sampling rate of 512 Hz. Horizontal and vertical eye movements were also recorded, and linked ears served as the reference lead. 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 kΩ. EEG epochs were synchronized with the onset of stimulus presentation. 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 μV and resulted in a rejection rate of ~5%. Evoked-response potentials (ERPs) from 100 ms before through 1000 ms after stimulus onset were averaged off-line. ERP components (including the site and latency to reach maximum amplitude) were identified and measured with respect to the baseline voltage, which was averaged over the interval from −100 ms to 0 ms.

The peak amplitude and latency of sensory P1 response was measured at mesial occipital (O1, O2) and lateral occipital (POO9h, POO10h) electrode sites, in the 80–120 ms time window. The mean amplitude of frontal N1 and N2 were measured at the left and right central (C1, C2, C3, C4), frontal (F1, F2, F3 and F4) and fronto-central (FC1, FC2, FC3 and FC4) electrode sites in the 100–120-ms and 200–275-ms time windows, respectively. The mean amplitude of the temporal P3 component was measured at the posterior temporal and temporo-parietal (T7, T8, TTP7h and TTP8h) electrode sites in the 600–800-ms time window. Multifactorial repeated measures were applied to the ERP data using the following within factors: stimulus category (Sound, Non-Sound), electrode (according to the ERP component of interest) and hemisphere (Left, Right). Multiple comparisons of means were performed by the post-hoc Tukey test. The alpha inflation due to multiple comparisons was corrected by means of Greenhouse-Geisser correction. The degrees of freedom accordingly modified are reported, together with ε and corrected probability level.

Low-Resolution Electromagnetic Tomography (LORETA) was performed on the ERP waveforms at the latency stage where the sound/non-sound difference was greatest, namely, at N1, N2, and P3 levels. LORETA[22] is a discrete linear solution to the inverse EEG problem and corresponds to the 3D distribution of neuronal electrical activity that has maximally similar (i.e., maximally synchronized) orientation and strength between neighboring neuronal populations (represented by adjacent voxels). In this study, an improved version of standardized weighted LORETA was used; this version, called swLORETA, incorporates a singular value decomposition-based lead field weighting method. The source space properties included grid spacing (the distance between two calculation points) of 5 points and an estimated signal-to-noise ratio (which defines the regularization; a higher value indicates less regularization and therefore less blurred results) of 3. SwLORETA was performed on the group data and identified statistically significant electromagnetic dipoles ($p < 0.05$) with larger magnitudes correlating with more significant activation. A realistic boundary element model (BEM) was derived from a T1-weighted 3D MRI data set by segmentation of the brain tissue. This BEM model consisted of one homogenous compartment comprised of 3,446 vertices and 6,888 triangles. The head model was used for intracranial localization of surface potentials. Both segmentation and generation of the head model were performed using the ASA software program.

1. Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H. & Martin, A. Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nat. Neurosci* **7**(11), 1190–1192 (2004).
2. Beauchamp, M. S., Lee, K. E., Argall, B. D. & Martin, A. Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* **41**(5), 809–823 (2004).
3. Tranel, D., Damasio, H., Eichhorn, G. R., Grabowski, T., Ponto, L. L., Hichwa, R. D. Neural correlates of naming animals from their characteristic sounds. *Neuropsychologia*, **41**(7), 847–854 (2003).
4. van Atteveldt, N., Formisano, E., Goebel, R. & Blomert, L. Integration of Letters and Speech Sounds in the Human Brain. *Neuron* **43**, 271 (2004).
5. Naue, N., Rach, S., Strüber, D., Huster, R. J., Zaehle, T., Körner, U., Herrmann, C. S. Auditory Event-Related Response in Visual Cortex Modulates Subsequent Visual Responses in Humans. *J. Neurosci.* **31**, 7729–7736 (2011).
6. Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O. V., Lu, S. T., Simola, J. Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neurosci. Lett.*, **127**(1), 141–145 (1991).
7. Kislyuk, D. S., Möttönen, R., Sams, M. Visual processing affects the neural basis of auditory discrimination. *J. Cogn. Neurosci.* **20**, 2175–2184 (2008).
8. Besle, J., Fischer, C., Bidet-Caulet, A., Lecaignard, F., Bertrand, O., Giard, M. H. Visual activation and audiovisual interactions in the auditory cortex during speech perception: intracranial recordings in humans. *J Neurosci*, **28**(52), 14301–14310 (2008).
9. Kohler, E., Keysers, C., Umiltà, M. A., Fogassi, L., Gallese, V., Rizzolatti, G. Hearing Sounds, Understanding Actions: Action Representation in Mirror Neurons. *Science* **297**, 846–848 (2002).
10. Keysers, C., Kohler, E., Umiltà, M. A., Nanetti, L., Fogassi, L., Gallese, V. Audiovisual mirror neurons and action recognition. *Exp. Brain Res.* **153**, 628 (2003).
11. Senkowski, D., Saint-Amour, D., Kelly, S. P. & Foxe, J. J. Multisensory processing of naturalistic objects in motion: a high-density electrical mapping and source estimation study. *Neuroimage* **36**(3), 877–888 (2007).
12. Besle, J., Bertrand, O. & Giard, M. H. Electrophysiological (EEG, sEEG, MEG) evidence for multiple audiovisual interactions in the human auditory cortex. *Hear Res.* **258**(1–2), 143–151 (2009).
13. Anderer, P., Saletu, B., Saletu-Zyhlarz, G., Gruber, D., Metka, M., Huber, J., Pascual-Marqui, R. D. Brain Regions Activated during an Auditory Discrimination Task in Insomniac Postmenopausal Patients before and after Hormone Replacement Therapy: Low-Resolution Brain Electromagnetic Tomography Applied to Event-Related Potentials. *Neuropsychobiology* **49**, 134 (2004).
14. Alpert, G. F., Hein, G., Tsai, N., Naumer, M. J. & Knight, R. T. Temporal Characteristics of Audiovisual Information Processing. *J. Neurosci.* **28**, 5344–5349 (2008).
15. Näätänen, R. & Winkler, I. The concept of auditory stimulus representation in cognitive neuroscience. *Psychol. Bull.* **125**(6), 826–859 (1999).
16. Wu, J., Yu, Z., Mai, X., Wei, J. & Luo, Y. Pitch and loudness information encoded in auditory imagery as revealed by event-related potentials. *Psychophysiology* **48**(3), 415–9 (2011).
17. Liebenthal, E., Desai, R., Ellingson, M. M., Ramachandran, B., Desai, A., Binder, J. R. Specialization along the Left Superior Temporal Sulcus for Auditory Categorization. *Cereb. Cortex* **20**, 2958–2970.
18. Beauchamp, M. S., Nath, A. R. & Pasalar, S. fMRI-Guided Transcranial Magnetic Stimulation Reveals That the Superior Temporal Sulcus Is a Cortical Locus of the McGurk Effect. *J. Neurosci.* **30**, 2414–2417.
19. Aglioti, S. & Pazzaglia, M. Representing actions through their sound. *Exp. Brain Res.* **206**, 141–151.
20. Bentaleb, L. A., Beauregard, M., Liddle, P. & Stip, E. Cerebral activity associated with auditory verbal hallucinations: a functional magnetic resonance imaging case study. *J Psychiatry Neurosci.* **27**(2), 110–5 (2002).
21. Levitan, C., Ward, P. B. & Catts, S. V. Superior temporal gyral volumes and laterality correlates of auditory hallucinations in schizophrenia. *Biol. Psychiatry* **46**, 955–62 (1999).
22. Pasqual-Marqui, R. D., Michel, C. M. & Lehmann, D. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *Int. J. Psychophysiol.* **18**, 49–65 (1994).

Acknowledgements

The authors are very grateful to Federica Riva, Mirella Manfredi and Nicola Crotti for their help in acquiring the EEG data. We gratefully acknowledge financial support from

University of Milano-Bicocca (2010 FAR funds) and IBFM-CNR. AR was supported in part by "Dote ricercatori": FSE, Regione Lombardia funding.

Author contributions

AMP and GED designed methods and experiments. AMP interpreted the results and wrote the paper. RA and GED performed data acquisition and analysis, AZ 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., D'Aniello, G.E., Adorni, R. & Zani, A. When a photograph can be heard: Vision activates the auditory cortex within 110 ms. *Sci. Rep.* **1**, 54; DOI:10.1038/srep00054 (2011).

SCIENTIFIC REPORTS

DOI: 10.1038/srep01122

SUBJECT AREAS:
AUDITORY SYSTEM
IMAGING
NEUROSCIENCE
BIOLOGICAL TECHNIQUES

SCIENTIFIC REPORTS:
1 : 54
DOI: 10.1038/srep00054
(2011)

Published:
4 August 2011

Updated:
17 January 2013

CORRIGENDUM: When a photograph can be heard: Vision activates the auditory cortex within 110 ms

Alice Mado Proverbio, Guido Edoardo D'Aniello, Roberta Adorni & Alberto Zani

Due to a technical error, the scale for the y-axis was reversed for Figures 1, 2 and 4. The corrected figures, along with their legends, appear below.

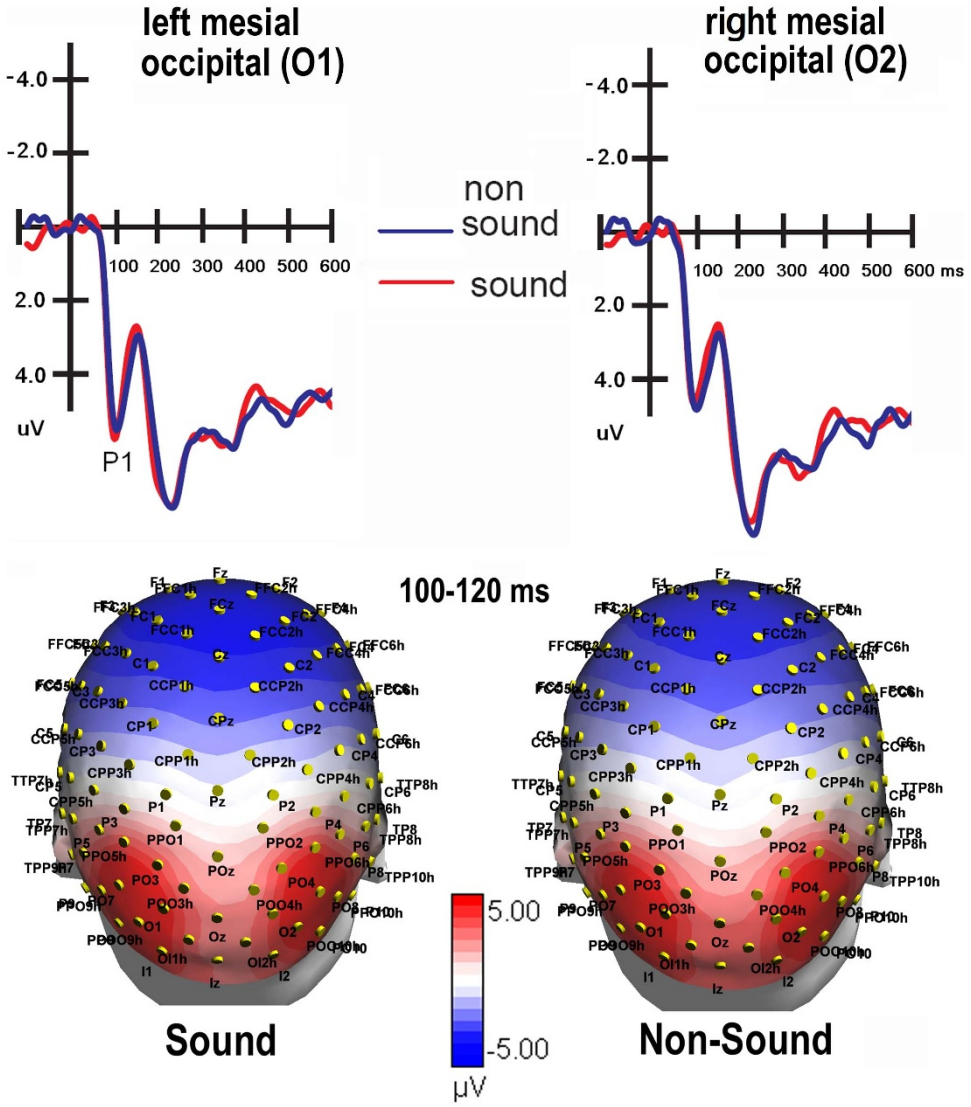

Figure 1 | (Top) Grand-average ERP waveforms recorded at left and right mesial occipital sites in response to sound and non-sound stimuli.

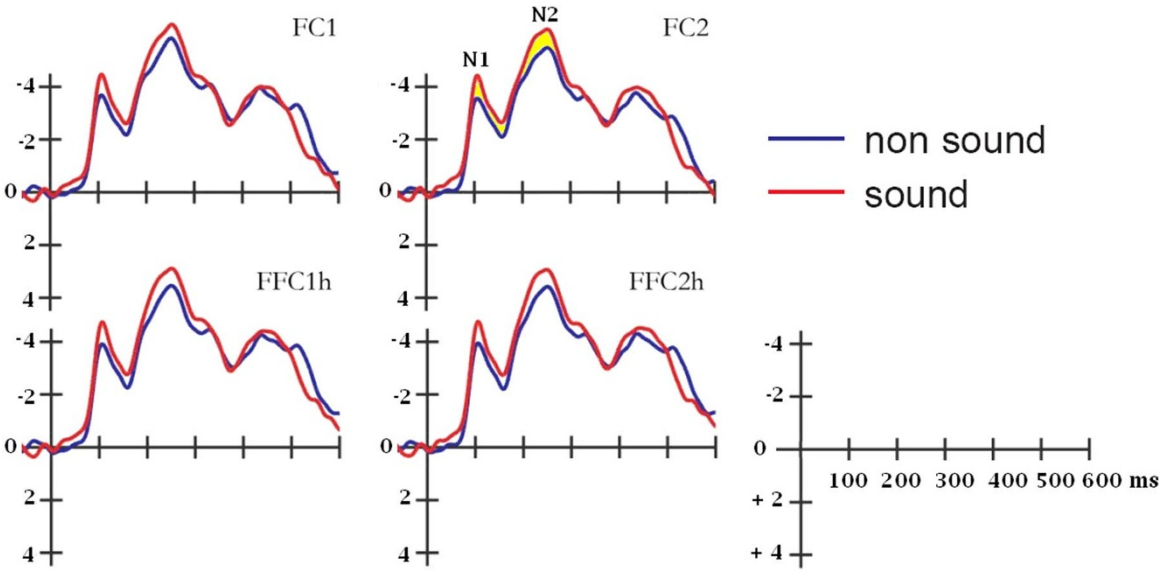

Figure 2 | Grand-average ERP waveforms recorded at left and right fronto-central sites in response to sound and non-sound stimuli.

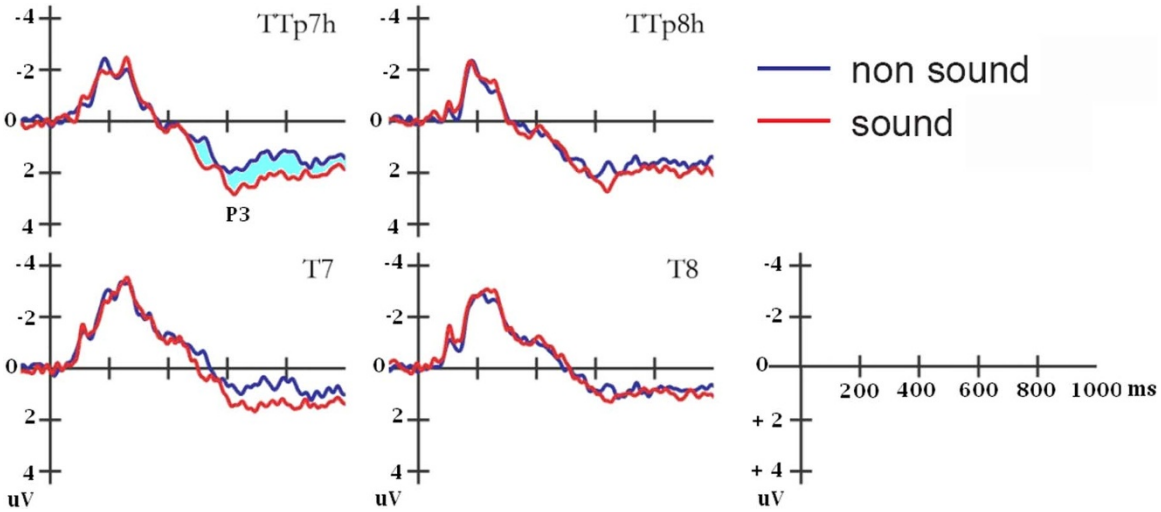

Figure 4 | Grand-average ERP waveforms recorded at left and right temporo-parietal and posterior-temporal sites in response to sound and non-sound stimuli.