

Onset timing of cross-sensory activations and multisensory interactions in auditory and visual sensory cortices

Tommi Raij,¹ Jyrki Ahveninen,¹ Fa-Hsuan Lin,^{1,2} Thomas Witzel,³ Iiro P. Jääskeläinen,⁴ Benjamin Letham,¹ Emily Israeli,³ Cherif Sahyoun,³ Christos Vasios,¹ Steven Stufflebeam,¹ Matti Hämäläinen^{1,3} and John W. Belliveau¹

¹MGH/MIT/HMS Athinoula A. Martinos Center for Biomedical Imaging, Bldg 149, 13th St, Charlestown, MA, USA

²Institute of Biomedical Engineering, National Taiwan University, Taipei, Taiwan

³Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA, USA

⁴Department of Biomedical Engineering and Computational Science, Aalto University, Espoo, Finland

Keywords: audiovisual interactions, auditory cortex, cross-modal, human, MEG, visual cortex

Abstract

Here we report early cross-sensory activations and audiovisual interactions at the visual and auditory cortices using magnetoencephalography (MEG) to obtain accurate timing information. Data from an identical fMRI experiment were employed to support MEG source localization results. Simple auditory and visual stimuli (300-ms noise bursts and checkerboards) were presented to seven healthy humans. MEG source analysis suggested generators in the auditory and visual sensory cortices for both within-modality and cross-sensory activations. fMRI cross-sensory activations were strong in the visual but almost absent in the auditory cortex; this discrepancy with MEG possibly reflects the influence of acoustical scanner noise in fMRI. In the primary auditory cortices (Heschl's gyrus) the onset of activity to auditory stimuli was observed at 23 ms in both hemispheres, and to visual stimuli at 82 ms in the left and at 75 ms in the right hemisphere. In the primary visual cortex (Calcarine fissure) the activations to visual stimuli started at 43 ms and to auditory stimuli at 53 ms. Cross-sensory activations thus started later than sensory-specific activations, by 55 ms in the auditory cortex and by 10 ms in the visual cortex, suggesting that the origins of the cross-sensory activations may be in the primary sensory cortices of the opposite modality, with conduction delays (from one sensory cortex to another) of 30–35 ms. Audiovisual interactions started at 85 ms in the left auditory, 80 ms in the right auditory and 74 ms in the visual cortex, i.e., 3–21 ms after inputs from the two modalities converged.

Introduction

Prevailing ideas on multisensory integration suggest that high-order heteromodal association cortical areas receive input from different sensory cortices and then integrate the signals (Mesulam, 1998). While there is much experimental evidence to support this notion (Cusick, 1997), recent research suggests that this view is incomplete. Specifically, low-order sensory areas may show cross-sensory (i.e., cross-modal) activations and multisensory interactions already starting at approximately 40–50 ms after the stimulus. The evidence includes intracranial electrophysiological recordings in nonhuman primates (Schroeder *et al.*, 2001; Schroeder & Foxe, 2002) and human electroencephalography (EEG) and magnetoencephalography (MEG) studies (Giard & Peronnet, 1999; Foxe *et al.*, 2000; Molholm *et al.*, 2002, 2004; Teder-Sälejärvi *et al.*, 2002; Murray *et al.*, 2005; Talsma *et al.*, 2007). Supporting evidence comes from functional magnetic resonance imaging (fMRI) studies that show such activations in or very close to primary sensory areas (Pekkola *et al.*, 2005; Martuzzi

et al., 2006). Taken together, these findings suggest that low-order sensory areas may contribute to multisensory integration starting from very early processing stages (Schroeder *et al.*, 2003; Foxe & Schroeder, 2005; Macaluso & Driver, 2005; Molholm & Foxe, 2005; Schroeder & Foxe, 2005; Ghazanfar & Schroeder, 2006; Macaluso, 2006; Musacchia & Schroeder, 2009).

Supporting these findings, anatomical connectivity studies in nonhuman primates have revealed direct cortico-cortical pathways from primary auditory (A1) to primary visual (V1) cortex (Rockland & Van Hoesen, 1994; Falchier *et al.*, 2002; Rockland & Ojima, 2003; Clavagnier *et al.*, 2004; Budinger *et al.*, 2006). Direct connections from V1 to A1 are not known, but the visual area V2 is directly connected with A1 (Budinger *et al.*, 2006).

Another slightly longer pathway between A1 and V1 is through the heteromodal association cortical area superior temporal polysensory area–superior temporal sulcus (STP/STS; Schroeder & Foxe, 2002; Cappe & Barone, 2005). This area also sends feedback to both V1 (Benevento *et al.*, 1977) and A1 (Smiley & Falchier, 2009 for a review). Further, STS is connected to the nonprimary supratemporal caudomedial auditory area, which has connections with A1 (de la Mothe *et al.*, 2006a).

Multisensory integration additionally takes place in several subcortical structures, of which the superior colliculus (SC) has been studied

Correspondence: Dr Tommi Raij, as above.
E-mail: rajj@nmr.mgh.harvard.edu

Some of these data have been presented in abstract form at the 8th annual meeting of the International Multisensory Research Forum (IMRF), Sydney, Australia, July 4–8, 2007.

Received 5 October 2009, revised 12 January 2010, accepted 15 January 2010

the most (Stein & Meredith, 1993). SC receives direct sensory input from central sensory pathways and then projects to multiple cortical areas (Stein & Meredith, 1993); it also receives cortical feedback (Stein *et al.*, 2002; Jiang & Stein, 2003). Connections between SC and STP/STS have been shown in primates (Bruce *et al.*, 1986; Gross, 1991). However, there are no known direct connections from SC to A1 or V1 (although SC receives direct input from V1, see Collins *et al.*, 2005). Anatomical connectivity mappings in nonhuman primates have revealed yet other possible subcortical and cortical locations from which A1 and V1 might receive multisensory inputs (de la Mothe *et al.*, 2006b; Hackett *et al.*, 2007; Smiley *et al.*, 2007; Cappe *et al.*, 2009; Musacchia & Schroeder, 2009; Smiley & Falchier, 2009). Here we examine the timing and possible pathways underlying early cross-sensory activations and audiovisual interactions by recording both MEG and fMRI responses in humans.

Materials and methods

Subjects, stimuli and tasks

Subjects were studied after they had given their written informed consent; the study protocol was approved by the Massachusetts General Hospital institutional review board and followed the guidelines of the Declaration of Helsinki. We presented 300-ms auditory (A), visual (V) and audiovisual (AV; simultaneous auditory and visual) stimuli to eight healthy right-handed human subjects (six females, age 22–30 years) in a rapid event-related fMRI design with pseudorandom stimulus order and interstimulus interval (ISI). A, V and AV stimuli were equiprobable. The A stimuli were white noise bursts (15 ms rise and decay) and the V stimuli static checkerboard patterns (visual angle $3.5^\circ \times 3.5^\circ$ and contrast 100%, foveal presentation). The task was to respond to rare (10%) target A (tone pips), V (checkerboard with a diamond pattern in the middle), or AV (combination) stimuli with the right index finger as quickly as possible, while the reaction time (RT) was measured. All subjects were recorded with three stimulus sequences with different ISIs. The three sequences had different mean (1.5, 3.1 and 6.1 s) ISIs; inside each sequence the ISI was jittered at 1.15 s (equivalent to TR of the fMRI acquisition) resolution to improve fMRI analysis power (Dale, 1999; Burock & Dale, 2000). All subjects were recorded with identical stimuli and tasks in both MEG and fMRI. The V stimuli were projected with a video projector onto a translucent screen. In MEG, the A stimuli were presented with MEG-compatible headphones. During fMRI the A stimuli were presented through MRI-compatible headphones (MR Confon GmbH, Magdeburg, Germany). Auditory stimulus intensity was adjusted to be as high as the subject could comfortably listen to (in MEG, approximately 65 dB SPL; in fMRI, clearly above the scanner acoustical noise). The stimuli were presented with a PC running Presentation 9.20 (Neurobehavioral Systems Inc, Albany, CA, USA). During fMRI the stimuli were synchronized with triggers from the fMRI scanner. The timing of the stimuli with respect to the trigger signals was confirmed with a digital oscilloscope.

Structural MRI recordings, brain segmentation and spatial intersubject alignment and morphing

Structural T1-weighted MRIs of the subjects were acquired with a 1.5T Siemens Avanto scanner (Siemens Medical Solutions, Erlangen, Germany) and a head coil using a standard MPRAGE sequence. Anatomical images were segmented with the FreeSurfer software (<http://www.surfer.nmr.mgh.harvard.edu>; Fischl *et al.*, 2002, 2004).

The individual brains were spatially co-registered by morphing them into the FreeSurfer average brain via a spherical surface (Fischl *et al.*, 1999).

fMRI recordings and analysis

Brain activity was measured using a 3.0T Siemens Trio scanner (Siemens Medical Solutions) with a Siemens head coil, and an echo planar imaging (EPI) sequence which is blood oxygenation level-dependent (BOLD; flip angle 90° , TR = 1.15 s, TE = 30 ms, 25 horizontal 4-mm slices with 0.4 mm gap, 3.1×3.1 mm in-plane resolution, fat saturation off). The rapid event-related functional data were analyzed with FreeSurfer (<http://www.surfer.nmr.mgh.harvard.edu>). During preprocessing, each individual's data were motion-corrected (Cox & Jesmanowicz, 1999), spatially smoothed with a Gaussian kernel of full-width at half maximum (FWHM) 5 mm, and normalized by scaling the whole brain intensity to a fixed value of 1000. The first three images of each run were discarded, as were rare images showing abrupt changes in intensity. Any remaining head motion was used as an external regressor. A finite impulse response (FIR) model (Burock & Dale, 2000) was applied to estimate the activations as a function of time separately for each trial type (A, V, AV, A Target, V Target, and AV Target) with a time window of 2.3 s prestimulus to 16.1 s poststimulus. The FIR method estimates the hemodynamic response time courses without assuming any form for the response. The functional volumes were spatially aligned with the structural MRI of individual subjects. During group analysis, the individual results were morphed through a spherical surface into the FreeSurfer average brain (Fischl *et al.*, 1999) and spatially smoothed at 10 mm FWHM.

MEG recordings

Whole-head 306-channel MEG (VectorView; Elekta-Neuromag, Finland) was recorded in a magnetically shielded room (Cohen *et al.*, 2002; Hämäläinen & Hari, 2002). The instrument employs three sensors (one magnetometer and two planar gradiometers) at each of the 102 measurement locations. We also recorded simultaneous horizontal and vertical electro-oculogram (EOG). All signals were bandpass-filtered to 0.03–200 Hz prior to sampling at 600 Hz.

Spatial registration of MEG data with MRI

Prior to the MEG recordings, the locations of four small head position indicator coils attached to the scalp and several additional scalp surface points were recorded with respect to the fiducial landmarks (nasion and two preauricular points) using a 3-D digitizer (Fastrak Polhemus, VT, USA). For MRI–MEG coordinate system alignment, the fiducial points were then identified from the structural MRIs. Using scalp surface locations, this initial approximation was refined using an iterative closest-point search algorithm.

MEG analysis of evoked responses

Responses were averaged offline separately for each trial type (A, V, AV, A Target, V Target, and AV Target) time-locked to the stimulus onsets with a time window of 250 ms prestimulus to 1150 ms poststimulus, with a total of 375 individual epochs per category for all nontarget conditions (100 epochs for the long, 125 for the intermediate, and 150 for the short ISI run). Epochs $> 150 \mu\text{V}$ or 3000 fT/cm at any EOG or MEG channel, respectively, were automatically discarded from the averages. For analysis of the MEG response

waveforms, the averaged signals were digitally lowpass-filtered at 40 Hz and amplitudes were measured with respect to a 200-ms prestimulus baseline. Nontarget A, V and AV evoked responses were analyzed for timing information, as were AV interactions estimated from the calculated response $[AV - (A + V)]$. For sensor analysis, we estimated the onset latencies from the gradient amplitudes $\sqrt{x^2 + y^2}$ from the two planar gradiometers x and y at each sensor location. Onset latencies were picked at the first time point that exceeded 3 SD above noise level estimated from the 200-ms prestimulus baseline. We additionally required that the onset must not occur earlier than 15 ms and the response had to stay above the noise level for at least 20 ms. Data from one subject were too noisy for accurate onset latency determination and were therefore discarded. Onset latencies from the three runs with different ISIs were practically identical; thus, the responses were averaged across ISI conditions, resulting in each subject's averaged response consisting of approximately 300 responses to individual stimuli (for detailed numbers of epochs see Supporting information, Appendix S1). Interaction responses had stronger noise (in sensor space, theoretically by $\sqrt{3}$ times) than their constituent (A, V, and AV) responses, requiring stronger lowpass filtering (20 Hz with 3 dB roll-off). Further, for the same reason the onsets picked from individual subjects' interaction responses were less reliable. We therefore used bootstrapping to estimate the means and variances of interaction onsets across subjects (for details see Supporting Information, Appendix S1).

MEG source analysis and source-specific time-course extraction

Minimum-norm estimates (MNEs; Hämäläinen & Ilmoniemi, 1984, 1994) were computed from combined anatomical MRI and MEG data (Dale & Sereno, 1993; Liu *et al.*, 1998; Dale *et al.*, 2000). The anatomically constrained MNE solutions were implemented in our software package available at <http://www.nmr.mgh.harvard.edu/martinos/userInfo/data/sofMNE.php>. For inverse computations, the cortical surface was decimated to 5000–10 000 vertices per hemisphere. A gain matrix **A** describing the ensemble of MEG sensor measurements with one current dipole on every vertex point was calculated using a realistic single-compartment boundary element model (Hämäläinen & Sarvas, 1989) based on the structural MRI data. The noise covariance matrix (**C**) was estimated from the prestimulus baselines of individual trials. These two matrices, along with the source covariance matrix **R**, were used to calculate the inverse operator $\mathbf{W} = \mathbf{R}\mathbf{A}^T(\mathbf{A}\mathbf{R}\mathbf{A}^T + \mathbf{C})^{-1}$. The MEG data at each time point were then multiplied by **W** to yield the estimated source activity in the cortical surface: $\mathbf{s}(t) = \mathbf{W}\mathbf{x}(t)$. Finally, dynamic statistical parametric mapping (dSPM) values (noise-normalized MNE) were calculated to reduce the point-spread function and to allow display of the activations using the *F*-statistic. Using the MNE software, the individual dSPM results were morphed through a spherical surface into the FreeSurfer average brain (Fischl *et al.*, 1999). Grand average dSPM estimates were calculated from the grand average MNE and the grand average noise covariance matrix. dSPM time courses were extracted from predetermined (Desikan *et al.*, 2006) anatomical locations of A1 and V1, after which their onset latencies were measured as described above for sensor signals.

Results

Behavioral results

Hit rates and RTs to target stimuli were measured during MEG and fMRI. Hit rates were excellent (96%) across all experimental

conditions. During MEG, the RTs were faster for AV (median 421 ms, mean \pm SD 430 \pm 60 ms) and A (median 462 ms, mean \pm SD 464 \pm 72 ms) than for V (median 520 ms, mean \pm SD 526 \pm 54 ms) stimuli with outliers excluded according to the median absolute deviation statistics criterion. In fMRI the difference was slightly smaller (for AV, median 539 ms, mean 546 \pm 86 ms; for A, median 584 ms, mean 596 \pm 112 ms; and for V stimuli, median 628 ms, mean 634 \pm 74 ms). The longer RTs in fMRI may be due to slower response pads and the MR environment. As the A, V and AV stimuli were in random order within stimulus sequences and stimulus timing was pseudorandom, it is unlikely that attention-related differences could have influenced the onset latencies. RT cumulative distributions showed behavioral evidence of multisensory integration (Raj T, unpublished observations).

MEG onset latencies: sensor data

Figure 1 shows that, as expected, the A stimuli activated the auditory cortices bilaterally and the V stimuli the midline occipital visual cortex. However, cross-sensory effects were also observed: V stimuli strongly activated temporal cortices and A stimuli (more weakly) the midline occipital cortex. Table 1 lists the corresponding onset latencies (the time when the grand average response first exceeded 3 SD noise level estimated from the prestimulus baseline). The sensory-specific activations started 19–22 ms earlier over the auditory than visual cortex. The cross-sensory activations started after the sensory-specific responses, by 21 ms over visual cortex and by 46 ms over auditory cortex. The left and right auditory cortices showed similar timings and were thus averaged for individual level analysis. Table 2 lists the across-subjects onset latencies. The individual subjects' responses were clearly noisier and thus relatively poorly corresponded to the grand average results; hence, no statistical comparisons were made for the sensor data (see dSPM data below for statistical tests).

MEG dSPM source analysis

Due to the relatively large distance between the sensors and the sources, each MEG sensor records activity from a rather large cortical area and MEG source analysis can better estimate the actual source locations. Figure 2 shows the MEG localization results at selected time points after the onset of activity. As expected, A stimuli activated the supratemporal auditory cortex and V stimuli the primary visual cortex in the calcarine fissure. Cross-sensory activations were also clear: V stimuli strongly activated large areas of temporal cortex including the supratemporal auditory cortex, and A stimuli (albeit more weakly) some parts of the calcarine fissure especially in the left hemisphere (right hemisphere cross-sensory activity in calcarine cortex was below selected visualization threshold). Additional cross-sensory activations were observed outside primary sensory areas.

MEG dSPM source-specific onset latencies

Figure 3 shows the source-specific grand average dSPM time courses from Heschl's gyri (auditory cortices) and calcarine fissure (visual cortex) for A and V stimuli. The areas were localized based on an anatomical parcellation of the FreeSurfer analysis package (Desikan *et al.*, 2006). The left and right calcarine fissure activations, due to their close anatomical proximity and similar timings and current orientations for foveal stimuli, were averaged. As expected, the dSPM

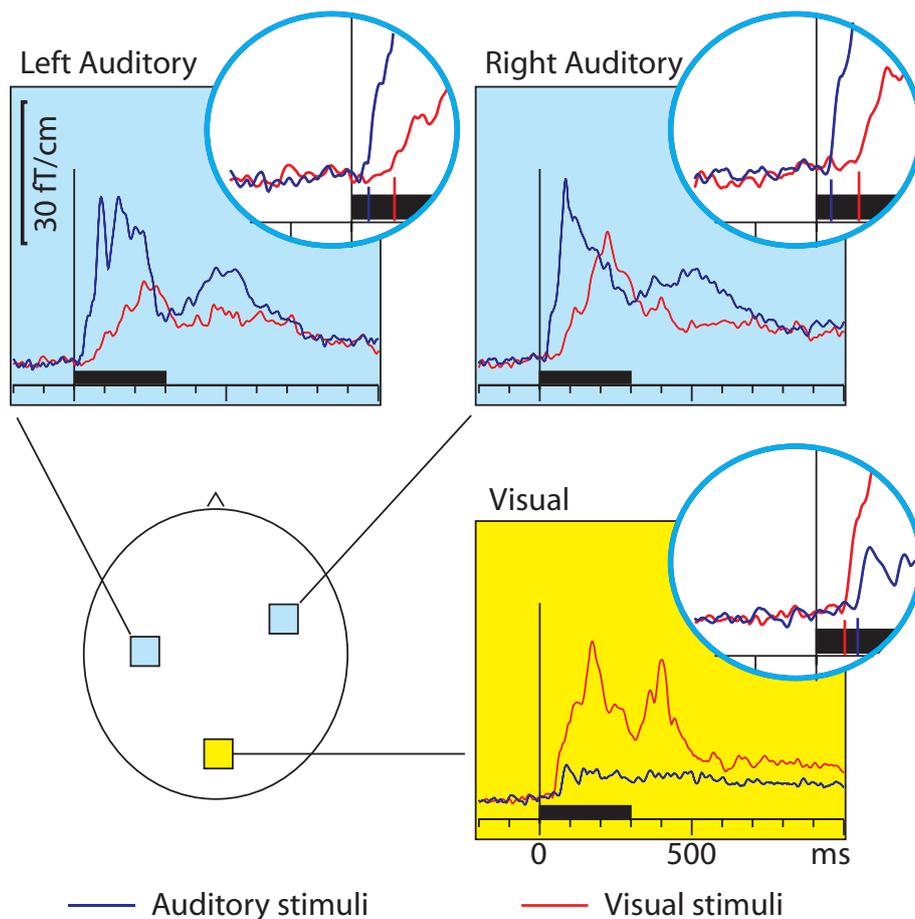


FIG. 1. MEG sensor responses over the auditory (light blue background) and visual (yellow background) cortices for auditory (blue traces) and visual (red traces) stimuli; the approximate sensor locations are shown in the lower left panel. The circular insets show the beginning of the response enlarged two-fold, with vertical lines where the onsets were found; the corresponding numerical values are reported in Table 1. The responses show the magnetic field gradient amplitudes as a function of time. From each subject, the sensor location showing the maximal approximately 100 ms sensory-specific response was selected, and the signals from these sensors were averaged across subjects. Sensors over both auditory and visual cortices showed cross-sensory activations, but these were stronger over the auditory than the visual cortex. The sensory-specific activations occurred earlier than the cross-sensory activations, especially over the auditory cortices. Time scales -200 to $+1000$ ms poststimulus, stimulus duration 300 ms (black bar).

TABLE 1. Grand average results in sensor space

	Sensor latencies (ms)		
	Auditory cortex (L)	Auditory cortex (R)	Visual cortex
Auditory stimuli	28	25	68
Visual stimuli	72	72	47
Audiovisual stimuli	25	27	40

The values were picked from the grand average sensor signals shown in Fig. 1; for across-subjects values see Table 2. L, left; R, right.

source-specific time courses were similar to those for the sensor data in Fig. 1. However, in the presence of multiple source areas, time courses extracted from specific cortical locations can more accurately reflect activity of the selected area than sensors that collect activity from a rather large area, due to volume conduction. Table 3 lists the onset latencies measured from the grand average dSPM responses. The sensory-specific activations started 20 ms earlier in A1 than in V1. The cross-sensory activations started after the sensory-specific responses, by 10 ms in V1 and by 59 ms in the left and 52 ms in the

TABLE 2. Across-subjects results in sensor space

	Sensor latencies (ms)	
	Auditory cortex (L & R)	Visual cortex
Auditory stimuli	30 ± 4	83 ± 40
Visual stimuli	84 ± 28	61 ± 12
Audiovisual stimuli	42 ± 9	50 ± 11

Mean \pm SD values.

right A1. The conduction delays (the time it takes for one stimulus to spread from one sensory cortex to the other) were 30 ms for A and 35 ms for V stimuli. As expected, the onsets of responses to AV stimuli (not shown in Fig. 3) closely followed the onsets to the unimodal stimulus that first reached the sensory cortex. Figure 4 shows the dSPM time courses calculated from the audiovisual interaction responses $[AV - (A + V)]$. These were much weaker than the constituent A, V, and AV responses and had a poorer signal-to-noise ratio (SNR). The interactions started 3–21 ms after inputs from the two sensory modalities converged on the sensory cortex.

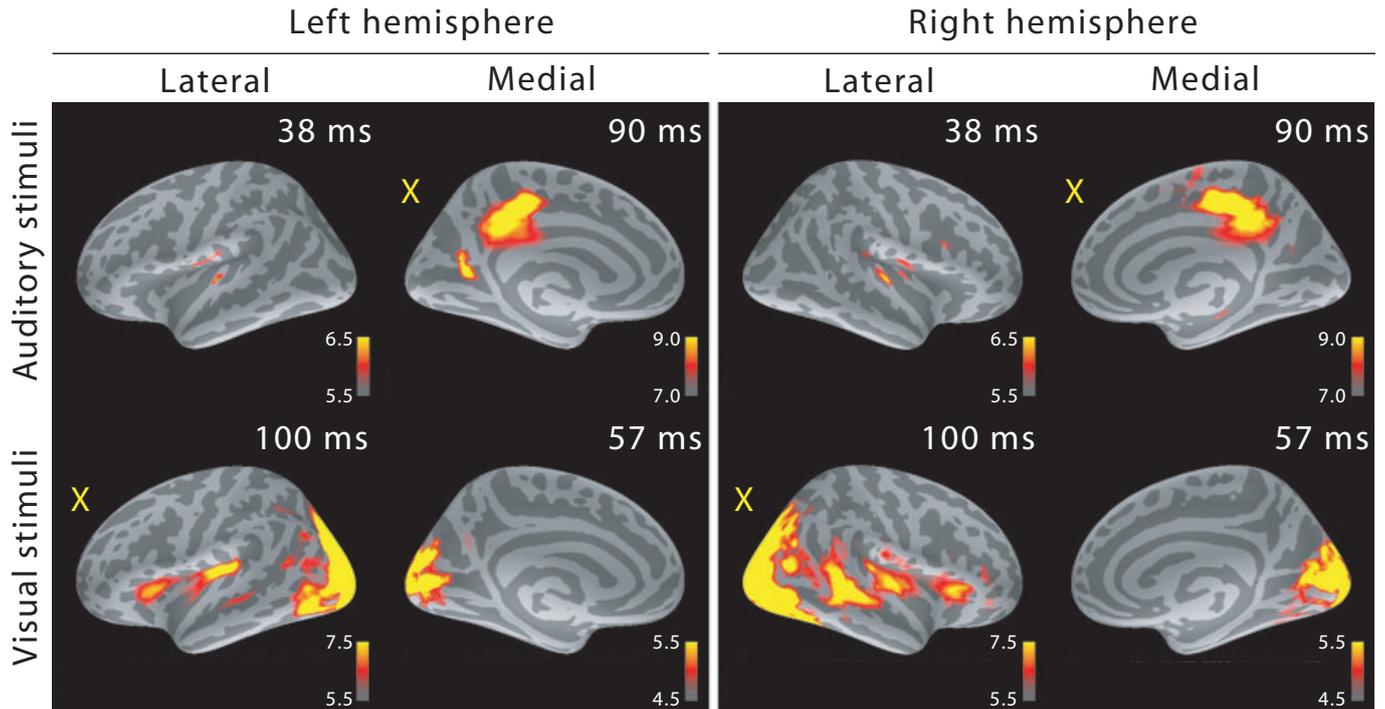


FIG. 2. MEG source analysis snapshots (dSPM F -statistics) picked at early activation latencies. Both sensory-specific and cross-sensory (marked with a yellow 'X') activations are seen (the right hemisphere calcarine cortex cross-sensory activity is not visible at this threshold). While some of the cross-sensory activations are located inside the sensory areas (as delineated in Desikan *et al.*, 2006), these seem to occupy slightly different locations than the sensory-specific activations. However, the spatial resolution of MEG is somewhat limited, hence exact comparisons are discouraged. Visual checkerboard stimuli activated additional areas outside the sensory cortices, for example superior temporal sulci (STS) especially in the right hemisphere and Broca's areas bilaterally.

The individual subjects' dSPM time courses had an improved SNR and much better corresponded to the latencies picked from the grand average responses than the sensor data (Fig. 1, Tables 1 and 2); hence, we considered the dSPM time courses the more accurate metrics. Again, as the onset latencies did not clearly differ across hemispheres, for individual level analysis the responses were averaged across the left and right hemisphere. Table 4 lists the across-subjects onset latencies (mean \pm SD ms and median across the latencies measured from the individual subjects' responses). The sensory-specific auditory evoked responses in Heschl's gyrus started 21 ms earlier than the visual evoked responses in the calcarine cortex (Wilcoxon signed rank test ($n = 7$), $P = 0.0156$). Cross-sensory activations in Heschl's gyrus occurred 49 ms later than sensory-specific activations, which was statistically significant ($P = 0.0156$). Cross-sensory activations in calcarine cortex occurred 22 ms later than sensory-specific activations, but this difference did not quite reach statistical significance ($P = 0.0781$). The difference between cross-sensory conduction delays (from one sensory cortex to another) for A and V stimuli was nonsignificant ($P = 0.578$). These nonparametric test results were highly consistent with confirmatory analyses conducted with parametric methods (paired t -tests).

fMRI activations

MEG source analysis has some uncertainty due to the electromagnetic inverse problem. We therefore attempted to confirm the MEG source analysis with fMRI using the same subjects and stimuli. Figure 5 shows the fMRI results averaged across subjects; the BOLD time courses are shown below the activation maps. The calcarine cortex was activated by both V and A stimuli. The auditory cortex showed

activity for A stimuli but, in contrast to the MEG results, for V stimuli showed only a tiny positive deflection at the typical BOLD signal peak latency (Fig. 5 time courses). At closer inspection some voxels in medial parts of Heschl's gyri were activated by V stimuli ($P < 0.01$ in the left and $P < 0.1$ in the right hemisphere; grand average fMRI signal, fixed-effects analysis) while the majority were not, diluting these effects in the spatial average across the entire region-of-interest.

Discussion

Here we report early cross-sensory activations and audiovisual interactions in both A1 and V1 in humans. The current study is to our knowledge the first to utilize both MEG and fMRI in the same subjects for this purpose, and has the advantage of offering spatiotemporally accurate estimates; individually, the methods offer compromises between spatial and temporal accuracy. The delay from sensory-specific to cross-sensory activity was 55 ms in the auditory and 10 ms in the visual cortex, which is clearly asymmetrical. This timing pattern reflects the fact that sensory-specific activations start earlier in the auditory (23 ms) than in the visual (43 ms) cortex, and is thus consistent with the idea that the origin of the cross-sensory activations is in the sensory cortex of the opposite stimulus modality, with approximately 30–35 ms conduction delay between the two areas. Audiovisual interactions were observed after both sensory-specific and cross-sensory inputs converged on the sensory cortex.

As MEG detects synchronous activity of thousands of neurons, the relationship between anatomical distance and conduction delay is not necessarily straightforward. Therefore, with the approximately 30 ms delay, the cross-sensory activations could utilize direct cortico-cortical connections between the auditory and visual cortices, connect through

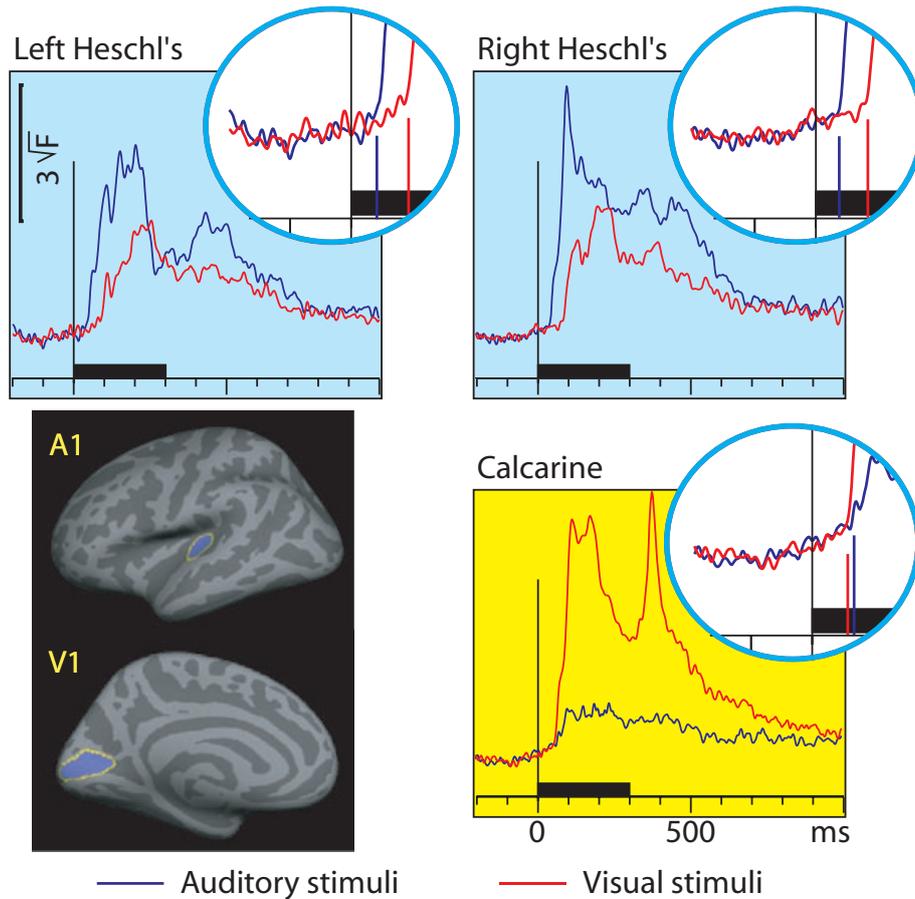


FIG. 3. MEG source-specific (dSPM) time courses for Heschl's gyri (auditory cortex; light blue background) and calcarine fissure (visual cortex; yellow background) to A and V stimuli; responses to AV stimuli are not shown. The source areas, shown for the left hemisphere in the lower left panel, were based on an anatomical parcellation (Desikan *et al.*, 2006); left and right calcarine sources were averaged. The circular insets show the beginning of each response enlarged two-fold, with vertical lines where the onsets for auditory (blue traces) and visual (red traces) stimuli were found; the corresponding numerical values are reported in Table 3. Both sensory-specific and cross-sensory activations are observed. The sensory-specific activations occurred earlier than the cross-sensory activations, especially in the auditory cortices. Time scales -200 to $+1000$ ms poststimulus, stimulus duration 300 ms (black bar).

TABLE 3. Grand average results in source space

	Source latencies (ms)		
	Heschl's gyrus (L)	Heschl's gyrus (R)	Calcarine cortex
Auditory stimuli	23	23	53
Visual stimuli	82	75	43
Audiovisual stimuli	23	27	47
Audiovisual interaction	85	80	74

The values were picked from dSPM grand average time courses shown in Figures 3–4; for across-subjects values see Table 4. L, left; R, right

a subcortical relay, or travel through an association cortical area such as STP/STS (e.g., Raij *et al.*, 2000). In the last option one would additionally expect activity in STS before observing cross-sensory activity in A1/V1. The analysis is complicated by the fact that, based on intracranial data from primates (Schroeder & Foxe, 2002; Schroeder *et al.*, 2003) and EEG recordings in humans (Foxe & Simpson, 2002), V stimuli would be expected to activate STS starting only approximately 8 ms after V1 onset, therefore largely overlapping

cross-sensory activations in the auditory cortex. In our data, STS was strongly activated in the right hemisphere at the same time as the cross-sensory auditory cortex activation occurred, consistent with the possibility of the signal traveling through STS, but in the left hemisphere no clear STS activation was observed. Hence, STS seems unlikely to play a key role. An additional factor to take into account is that the conduction delay had a small asymmetric trend: 30 ms for A stimuli with a monosynaptic connection $A1 \rightarrow V1$ and 35 ms for V stimuli with a known somewhat longer known pathway $V1 \rightarrow V2 \rightarrow A1$. Hence, it appears plausible that the earliest cross-sensory activations may utilize the $A1 \rightarrow V1$ and $V1 \rightarrow V2 \rightarrow A1$ pathways. Future studies utilizing dynamic causality modeling (Lin *et al.*, 2009; Schoffelen & Gross, 2009) might provide additional insight.

As described in Introduction, another possibility is that subcortical pathways may send direct cross-sensory inputs to sensory cortices. If the subcortical structures have a similar delay between auditory and visual processing as A1 and V1, then latency data alone cannot distinguish between cortico-cortical and subcortico-cortical cross-sensory influences. However, currently no such audiovisual pathways are known. Clearly, correct interpretation of functional connectivity analyses greatly benefits from accurate anatomical connectivity information.

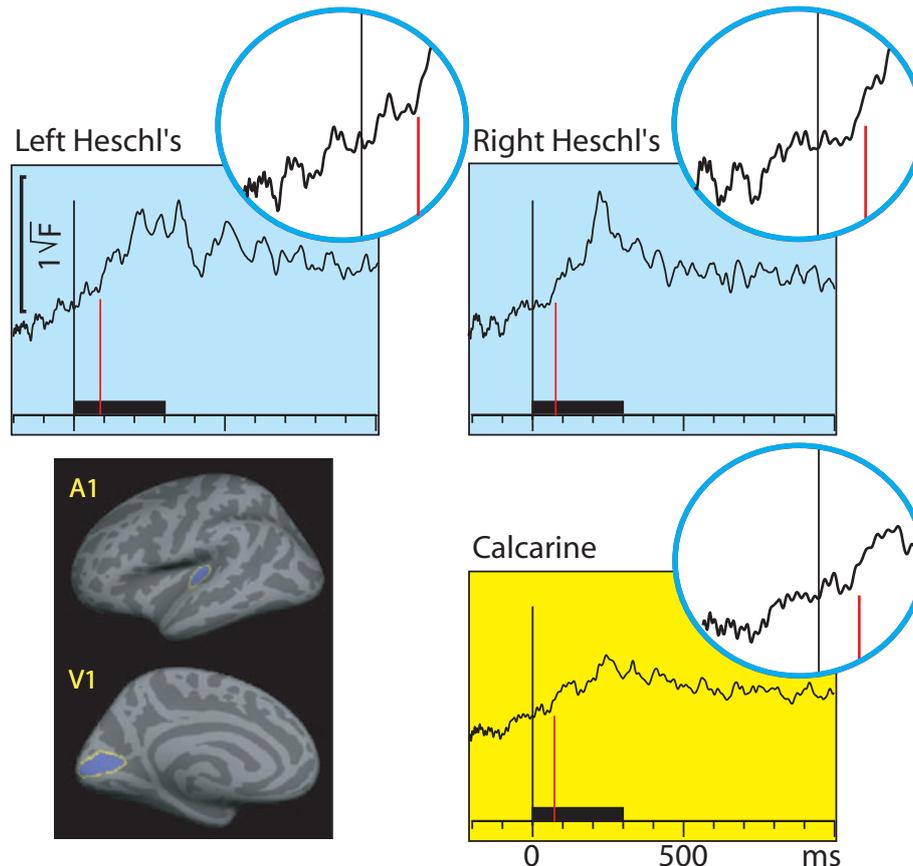


FIG. 4. MEG source-specific (dSPM) audiovisual interaction $[AV - (A + V)]$ time courses from Heschl's gyri (auditory cortex; light blue background) and calcarine fissure (visual cortex; yellow background). The circular insets show the beginning of each response enlarged two-fold, with vertical lines where the onsets were found; the corresponding numerical values are reported in Table 3. Interactions were observed in both the auditory and visual cortices, starting 3–21 ms after the inputs from the two sensory modalities converged in the sensory cortex. Time scales -200 to $+1000$ ms poststimulus, stimulus duration 300 ms (black bar).

TABLE 4. Across-subjects results in source space

	Source latencies (ms)	
	Heschl's gyrus (L & R)	Calcarine cortex
Auditory stimuli	27 ± 8 (28)	70 ± 26 (73)
Visual stimuli	76 ± 10 (77)	48 ± 8 (45)
Audiovisual stimuli	29 ± 11 (27)	50 ± 6 (52)
Audiovisual interaction*	77 ± 22 (84)	68 ± 28 (75)

Mean \pm SD values (Median in parentheses). *Interaction responses were low-pass filtered at 20 Hz and their mean, SD and median estimated by bootstrapping to mitigate SNR problems caused by the $[AV - (A + V)]$ operation. To see onset latencies for all stimulus categories using bootstrapping, see supporting Table S1.

The current results could mistakenly be interpreted to suggest that earliest audiovisual interactions can occur only after the cross-sensory inputs arrive at the sensory cortex. This would put a lower limit of 53 ms in the visual cortex and 75 ms for auditory cortex for audiovisual interactions to start, which is in fact what was observed in the present MEG data, yet there is strong EEG evidence of audiovisual interactions in humans occurring earlier, starting at approximately 40 ms, being maximal over posterior areas (Giard & Peronnet, 1999; Molholm *et al.*, 2002, 2004; Teder-Sälejärvi *et al.*, 2002). We suggest three possibilities as to why these early interactions

were not observed in the present MEG study. First, EEG may receive somewhat stronger contribution from subcortical generators than MEG (Goldenholz *et al.*, 2009), which is consistent with the idea that the early interactions in EEG may be generated in subcortical structures participating in multisensory processes. Second, the subcortical parts of afferent pathways leading to sensory cortex could be modulated by subcortical multisensory influences, which would allow audiovisual interactions to occur from the very beginning of the cortically generated 'sensory-specific' responses. However, this scenario would predict that the early interactions should be equally visible for EEG and MEG. Third, due to the sensitivity of MEG to mainly tangentially oriented currents, we could have missed some earlier components if they were radial. However, this is unlikely given that it has been estimated that only approximately 10% of the cortical surface (thin strips at crests of gyri) are radial enough to generate currents undetectable with MEG (Hillebrand & Barnes, 2002) and, further, source orientation differences would be expected to influence all activations and interactions equally because in the present study source areas were kept constant. Therefore, the most likely explanation is that the early interactions are generated in subcortical structures. EEG and MEG source localization accuracy for deep generators is poor, resulting in that these methods are not well suited for more accurate localization of the subcortical structures.

The finding that fMRI could detect strong cross-sensory activations in the calcarine fissure but in Heschl's gyri these were almost absent was unexpected. In the present data some voxels in Heschl's gyri were

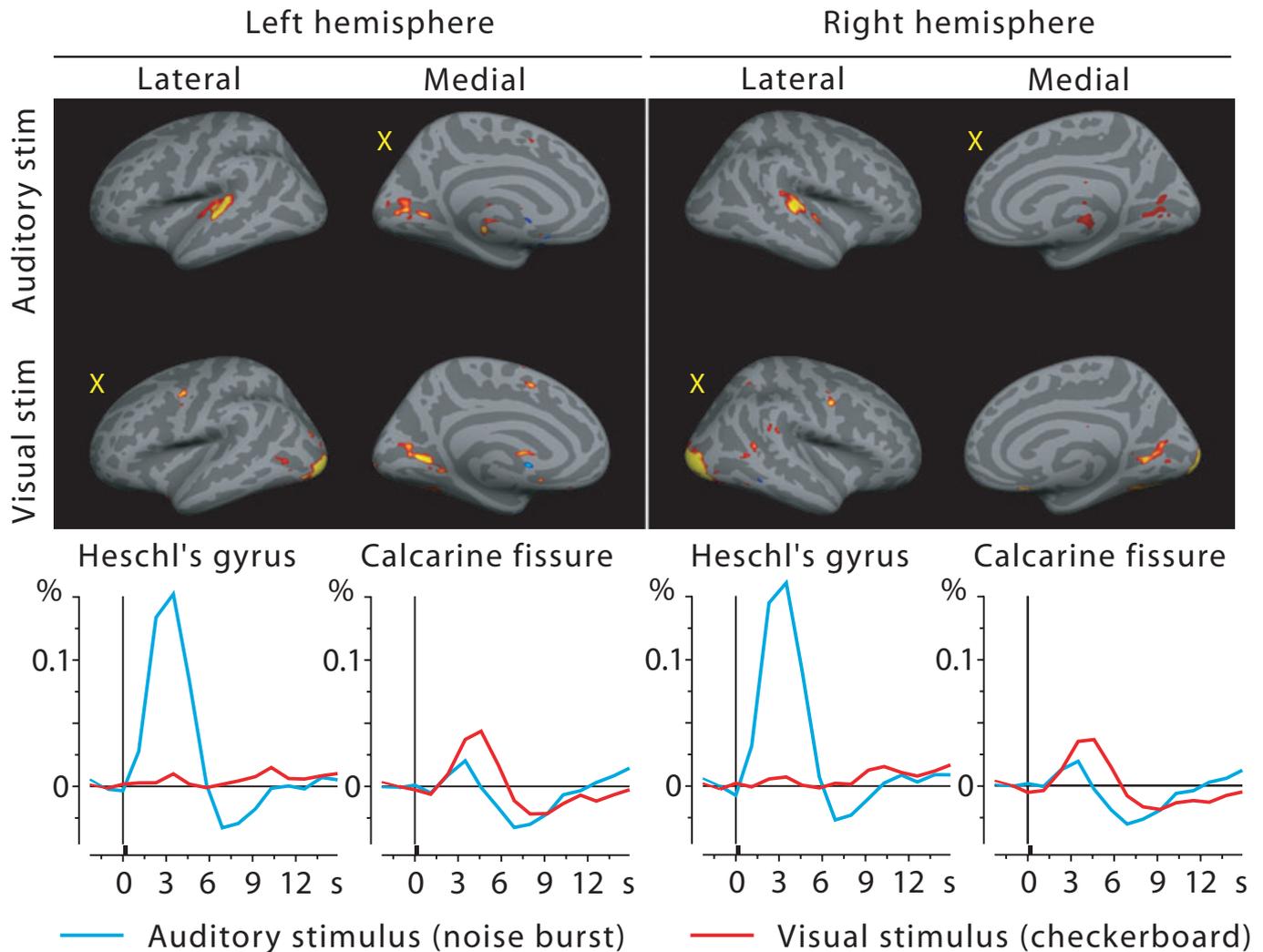


Fig. 5. fMRI activations to A and V stimuli projected on the inflated cortex at the fourth time frame after stimulus onset (top) and the corresponding BOLD % signal change time courses from Heschl's gyrus and calcarine fissure (bottom). Sensory-specific activations were very clear; cross-sensory responses were strong in the calcarine fissure but almost absent in Heschl's gyri (see Discussion). Yellow 'X' letters in the brain images mark cross-sensory conditions. Responses to AV stimuli and audiovisual interactions not shown.

significantly activated by V stimuli (albeit weakly) at the typical BOLD signal peak latency (see Fig. 5) while the majority were not, rendering the reliability of this observation inconclusive. Previous fMRI studies have shown that at least some classes of V stimuli (such as lip movements) may robustly activate A1 (Pekkola *et al.*, 2005). Even simple stimuli such as those employed in the current study have been reported to result in cross-sensory activations (Martuzzi *et al.*, 2007). One possible explanation is that the acoustical EPI scanner noise dampened responses in the auditory cortex due to neuronal adaptation. It is also possible that, again, due to the acoustical scanner noise, the BOLD signal may saturate before the neurons do (Bandettini *et al.*, 1998). A possible reason why our study may have been affected by this more than the above-mentioned could be that the acoustic noise is EPI parameter-dependent: our faster scanning could have increased the noise. This interpretation is supported by the observation that MEG, where the scanner is completely quiet, showed clear cross-sensory responses in the supratemporal auditory cortex.

It is unclear what the functional roles of the early cross-sensory activations might be. Behaviorally, for complex processing such as audiovisual speech, asynchrony as large as 250 ms can go unnoticed (Miller & D'Esposito, 2005). Moreover, in realistic stimulus environ-

ments auditory input lags the visual input, depending on the distance from the source (9 ms increase for every 3 m distance), which influences the relative timings of the auditory and visual inputs. Possibly the early cross-sensory influences have a role for lower-order processing (where synchrony requirements may be tighter) than audiovisual speech. There is also evidence that these activations may be task-dependent (Wang *et al.*, 2008). Plausibly, early cross-sensory activations could serve to facilitate later processing stages and reaction times by enhancing top-down processing and speeding up the exchange of signals between brain areas (Bar *et al.*, 2006; Raij *et al.*, 2008; Sperdin *et al.*, 2009).

As a technical finding, a very high SNR was necessary in order to detect onset latencies accurately. The present results were achieved by using a low-noise MEG instrument, a high-quality shielded room and a large number of stimuli. The averaged responses in the current study consisted of approximately 300 individual responses per subject, which was not quite sufficient for sensor space analysis at the individual subject level but quite sufficient for grand average analysis (approximately 2100 individual responses, or twice as much when additionally averaging across hemispheres). However, compared with the sensor data, extracting time courses from the auditory and visual

sensory cortices by dSPM source analysis greatly improved SNR at the individual level (more robust onsets and less interindividual variability), hence giving more accurate results that also agreed with the grand average values well. Moreover, in both sensor and source space, we present two different across-subjects analyses: onset latencies picked (i) from grand average ($N = 7$) responses (Tables 1 and 3) and (ii) from individual subjects' responses (Tables 2 and 4). The latter were useful for testing the statistical significance of latency differences across areas. However, the grand average response consists of the largest number of epochs and therefore has by far the best SNR, consequently showing slightly earlier onsets than those picked from the individual subjects' responses (e.g., compare Tables 3 and 4). Yet, grand average responses could also be biased to show early onsets if some of the subjects have earlier onsets than the others. In our data this bias appears to be quite small as most latencies are similar across Tables 3 and 4. Still, differentiating between the boost given by improved SNR and the possible bias caused by subjects with faster onsets is difficult. The two analyses complement each other and offer slightly different interpretations. The grand average analysis is well suited for finding the earliest onset latencies across the subject pool. The means across values from individual subjects, in turn, show slightly longer onset latencies but are better protected from individual bias. An additional possibility would be to use bootstrapping to synthesize multiple grand averages and, after picking onsets from each, study their means and variances. The results, shown in supporting Table S1, may offer a compromise between the two analyses. With the present data all three analyses give quite similar results and lead to the same conclusions; to improve comparability with earlier studies we here focus on reporting the results with the most widely used methods.

The current results are not directly comparable with studies where stimuli or tasks in one modality precede the other. For example, in audiovisual speech, the visual input (lip movements) typically starts 100–300 ms before the auditory stimulus onset and therefore may modulate the incoming auditory signals at multiple levels, including in secondary auditory cortex (Besle *et al.*, 2008) and even in central auditory pathways (Musacchia *et al.*, 2006). Similarly, auditory evoked responses can already be modulated by visuomotor processes such as gaze direction in the inferior colliculus (Groh *et al.*, 2001). As yet another example, attention may modulate responses and interactions through top-down mechanisms in primary sensory cortices as soon as they appear (Talsma *et al.*, 2007; Poghosyan & Ioannides, 2008; Karns & Knight, 2009). The flash-sound illusion also would appear to belong in this category (Shams *et al.*, 2002, 2005; Watkins *et al.*, 2006; Mishra *et al.*, 2007).

In the current study V stimuli were presented foveally. However, anatomical studies have shown that areas in the calcarine fissure representing peripheral vision may be more strongly connected with the auditory cortex than areas representing the fovea (Falchier *et al.*, 2002; Wang *et al.*, 2008). It is therefore plausible that cross-sensory latencies could be faster for peripherally than for foveally presented visual stimuli, although some previous studies have found the opposite effect (Talsma & Woldorff, 2005; Talsma *et al.*, 2007).

The late BOLD negative undershoots for A stimuli in the calcarine cortex (Fig. 5) are consistent with an earlier block design fMRI study reporting cross-sensory negative BOLD activations (Laurienti *et al.*, 2002); however, due to their study design, they could not investigate the time courses of the BOLD responses. Our BOLD time course analysis shows that the cross-sensory responses in the visual cortex show a small initial positive component, followed by a clearly stronger negative deactivation component. Temporal summation of such events in a block design would be expected to result in a net negative BOLD effect.

These findings are consistent with previously shown sensory-specific and cross-sensory activations (see Introduction). For example, the A1 onsets for our A stimuli at 23 ms are only approximately 8 ms slower than the earliest responses to clicks recorded from the human auditory cortex intracranially (Celesia, 1976) or by MEG (Parkkonen *et al.*, 2009), and the V1 onset at 43 ms simultaneous with the earliest reported responses from V1 (Foxe & Schroeder, 2005; Musacchia & Schroeder, 2009 for reviews). The observed cross-sensory onset latencies are, to our knowledge, the fastest reported in humans. This was made possible by the good SNR in our data and the extraction of source-specific amplitudes. Audiovisual interactions were observed only after the uni- and cross-sensory inputs converged on the sensory cortex, but once this happened the interactions appeared almost instantaneously (3–21 ms after convergence). The findings contribute to understanding of cross-sensory activations and interactions in sensory cortices by establishing lower limits to the latency when they can be expected to occur. The results have implications regarding the possible pathways that cross-sensory activations utilize, and suggest that audiovisual interactions occurring before cross-sensory signals arrive (for simultaneous stimuli, 53 ms in visual cortex and 75 ms in auditory cortex) are most probably of subcortical origin; interactions after these latencies could be either cortically or subcortically generated.

Supporting Information

Additional supporting information may be found in the online version of this article:

Table S1. Cross-subjects results in source space using bootstrapping. Appendix S1. Supporting materials and methods.

Please note: As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset by Wiley-Blackwell. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Acknowledgements

We would like to thank Valerie Carr, Deirdre Foxe, Mark Halko, Hsiao-Wen Huang, Yu-Hua Huang, Adrian KC Lee, Natsuko Mori, Mark Vangel and Dan Wakeman for invaluable help. This work was supported by grants from the National Institutes of Health R01 NS048279, R01 HD040712, R01 NS037462, P41 RR14075 and R21EB007298, the National Center for Research Resources, the Harvard Catalyst Pilot Grant / The Harvard Clinical and Translational Science Center (NIH UL1 RR 025758-02 and financial contributions from participating organizations), the Sigrid Juselius Foundation, the Academy of Finland, the Finnish Cultural Foundation, the National Science Council, Taiwan (NSC 98-2320-B-002-004-MY3, NSC 97-2320-B-002-058-MY3) and the National Health Research Institute, Taiwan (NHRI-EX97-9715EC).

Abbreviations

A, auditory (stimulus); A1, primary auditory cortex; AV, audiovisual (stimulus); BOLD, blood oxygen level-dependent; dSPM, dynamic statistical parametric mapping; EEG, electroencephalography; EOG, electro-oculogram; EPI, echo planar imaging; fMRI, functional magnetic resonance imaging; ISI, interstimulus interval; MEG, magnetoencephalography; MNE, minimum-norm estimate; RT, reaction time; SC, superior colliculus; SNR, signal-to-noise ratio; STP/STS, superior temporal polysensory area–superior temporal sulcus; V, visual (stimulus); V1, primary visual cortex.

References

- Bandettini, P., Jesmanowicz, A., Van Kylen, J., Birn, R. & Hyde, J. (1998) Functional MRI of brain activation induced by scanner acoustic noise. *Magn. Reson. Med.*, **39**, 410–416.

- Bar, M., Kassam, K., Ghuman, A., Boshyan, J., Schmid, A., Dale, A., Hämäläinen, M., Marinkovic, K., Schacter, D., Rosen, B. & Halgren, E. (2006) Top-down facilitation of visual recognition. *Proc. Natl Acad. Sci. USA*, **103**, 449–454.
- Benevento, L., Fallon, J., Davis, B. & Rezak, M. (1977) Auditory-visual interaction in single cells in the cortex of the superior temporal sulcus and orbital cortex of the macaque monkey. *Exp. Neurol.*, **57**, 849–872.
- Besle, J., Fischer, C., Bidet-Caulet, A., Lecaigard, F., Bertrand, O. & Giard, M. (2008) Visual activation and audiovisual interactions in the auditory cortex during speech perception: intracranial recordings in humans. *J. Neurosci.*, **28**, 14301–14310.
- Bruce, C., Desimone, R. & Gross, C. (1986) Both striate cortex and superior colliculus contribute to visual properties of neurons in superior temporal polysensory area of macaque monkey. *J. Neurophysiol.*, **55**, 1057–1075.
- Budinger, E., Heil, P., Hess, A. & Scheich, H. (2006) Multisensory processing via early cortical stages: connections of the primary auditory cortical field with other sensory systems. *Neuroscience*, **143**, 1065–1083.
- Burock, M. & Dale, A. (2000) Estimation and detection of event-related fMRI signals with temporally correlated noise: a statistically efficient and unbiased approach. *Hum. Brain Mapp.*, **11**, 249–260.
- Cappe, C. & Barone, P. (2005) Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. *Eur. J. Neurosci.*, **22**, 2886–2902.
- Cappe, C., Morel, A., Barone, P. & Rouiller, E. (2009) The thalamocortical projection systems in primate: an anatomical support for multisensory and sensorimotor interplay. *Cereb. Cortex*, **19**, 2025–2037.
- Celesia, G. (1976) Organization of auditory cortical areas in man. *Brain*, **99**, 403–414.
- Clavagnier, S., Falchier, A. & Kennedy, H. (2004) Long-distance feedback projections to area V1: implications for multisensory integration, spatial awareness, and visual consciousness. *Cogn. Affect. Behav. Neurosci.*, **4**, 117–126.
- Cohen, D., Schlapfer, U., Ahlfors, S., Hamalainen, M. & Halgren, E. (2002) New six-layer magnetically-shielded room for MEG. In Nowak, H., Haueisen, J., Giessler, F. & Huonker, R. (Eds), *13th International Conference on Biomagne*. VDE Verlag, Jena, Germany, pp. 919–921.
- Collins, C., Lyon, D. & Kaas, J. (2005) Distribution across cortical areas of neurons projecting to the superior colliculus in new world monkeys. *Anat. Rec. A Discov. Mol. Cell. Evol. Biol.*, **285**, 619–627.
- Cox, R. & Jesmanowicz, A. (1999) Real-time 3D image registration for functional MRI. *Magn. Reson. Med.*, **42**, 1014–1018.
- Cusick, C. (1997) The superior temporal polysensory region in monkeys. In Rockland, K., Kaas, J. & Peters, A. (Eds), *Cerebral Cortex Volume 12: Extrastriate Cortex in Primates*. Plenum Press, New York, pp. 435–468.
- Dale, A. (1999) Optimal experimental design for event-related fMRI. *Hum. Brain Mapp.*, **8**, 109–114.
- Dale, A. & Sereno, M. (1993) Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach. *J. Cogn. Neurosci.*, **5**, 162–176.
- Dale, A., Liu, A., Fischl, B., Buckner, R., Belliveau, J., Lewine, J. & Halgren, E. (2000) Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron*, **26**, 55–67.
- Desikan, R., Segonne, F., Fischl, B., Quinn, B., Dickerson, B., Blacker, D., Buckner, R., Dale, A., Maguire, R., Hyman, B., Albert, M. & Killiany, R. (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*, **31**, 968–980.
- Falchier, A., Clavagnier, S., Barone, P. & Kennedy, H. (2002) Anatomical evidence of multimodal integration in primate striate cortex. *J. Neurosci.*, **22**, 5749–5759.
- Fischl, B., Sereno, M., Tootell, R. & Dale, A. (1999) High-resolution inter-subject averaging and a coordinate system for the cortical surface. *Hum. Brain Mapp.*, **8**, 272–284.
- Fischl, B., Salat, D., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B. & Dale, A. (2002) Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, **33**, 341–355.
- Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Segonne, F., Salat, D., Busa, E., Seidman, L., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B. & Dale, A. (2004) Automatically parcellating the human cerebral cortex. *Cereb. Cortex*, **14**, 11–22.
- Foxe, J. & Schroeder, C. (2005) The case for feedforward multisensory convergence during early cortical processing. *Neuroreport*, **16**, 419–423.
- Foxe, J. & Simpson, G. (2002) Flow of activation from V1 to frontal cortex in humans. A framework for defining “early” visual processing. *Exp. Brain Res.*, **142**, 139–150.
- Foxe, J., Morocz, I., Murray, M., Higgins, B., Javitt, D. & Schroeder, C. (2000) Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Brain Res. Cogn. Brain Res.*, **10**, 77–83.
- Ghazanfar, A. & Schroeder, C. (2006) Is neocortex essentially multisensory? *Trends Cogn. Sci.*, **10**, 278–285.
- Giard, M. & Peronnet, F. (1999) Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J. Cogn. Neurosci.*, **11**, 473–490.
- Goldenholz, D., Ahlfors, S., Hämäläinen, M., Sharon, D., Ishitobi, M., Vaina, L. & Stufflebeam, S. (2009) Mapping the signal-to-noise-ratios of cortical sources in magnetoencephalography and electroencephalography. *Hum. Brain Mapp.*, **30**, 1077–1086.
- Groh, J., Trause, A., Underhill, A., Clark, K. & Inati, S. (2001) Eye position influences auditory responses in primate inferior colliculus. *Neuron*, **29**, 509–518.
- Gross, C. (1991) Contribution of striate cortex and the superior colliculus to visual function in area MT, the superior temporal polysensory area and the inferior temporal cortex. *Neuropsychologia*, **29**, 497–515.
- Hackett, T., De La Mothe, L., Ulbert, I., Karmos, G., Smiley, J. & Schroeder, C. (2007) Multisensory convergence in auditory cortex, II. Thalamocortical connections of the caudal superior temporal plane. *J. Comp. Neurol.*, **502**, 924–952.
- Hämäläinen, M. & Hari, R. (2002) Magnetoencephalographic characterization of dynamic brain activation. Basic principles and methods of data collection and source analysis. In Toga, A.W. & Mazziotta, J.C. (Ed.), *Brain Mapping: the Methods*. Academic Press, New York, pp. 227–253.
- Hämäläinen, M. & Ilmoniemi, R. (1984) *Interpreting Measured Magnetic Fields of the Brain: Estimates of Current Distributions*. Helsinki University of Technology, Helsinki, Finland.
- Hämäläinen, M. & Ilmoniemi, R. (1994) Interpreting magnetic fields of the brain: minimum norm estimates. *Med. Biol. Eng. Comput.*, **32**, 35–42.
- Hämäläinen, M. & Sarvas, J. (1989) Realistic conductivity geometry model of the human head for interpretation of neuromagnetic data. *IEEE Trans. Biomed. Eng.*, **36**, 165–171.
- Hillebrand, A. & Barnes, G. (2002) A quantitative assessment of the sensitivity of whole-head MEG to activity in the adult human cortex. *Neuroimage*, **16**, 638–650.
- Jiang, W. & Stein, B. (2003) Cortex controls multisensory depression in superior colliculus. *J. Neurophysiol.*, **90**, 2132–2135.
- Karns, C. & Knight, R. (2009) Intermodal auditory, visual, and tactile attention modulates early stages of neural processing. *J. Cogn. Neurosci.*, **21**, 669–683.
- Laurienti, P., Burdette, J., Wallace, M., Yen, Y., Field, A. & Stein, B. (2002) Deactivation of sensory-specific cortex by cross-modal stimuli. *J. Cogn. Neurosci.*, **14**, 420–429.
- Lin, F., Hara, K., Solo, V., Vangel, M., Belliveau, J., Stufflebeam, S. & Hämäläinen, M. (2009) Dynamic Granger-Geweke causality modeling with application to interictal spike propagation. *Hum. Brain Mapp.*, **30**, 1877–1886.
- Liu, A., Belliveau, J. & Dale, A. (1998) Spatiotemporal imaging of human brain activity using functional MRI constrained magnetoencephalography data: Monte Carlo simulations. *Proc. Natl. Acad. Sci. USA*, **95**, 8945–8950.
- Macaluso, E. (2006) Multisensory processing in sensory-specific cortical areas. *Neuroscientist*, **12**, 327–338.
- Macaluso, E. & Driver, J. (2005) Multisensory spatial interactions: a window onto functional integration in the human brain. *Trends Neurosci.*, **28**, 264–271.
- Martuzzi, R., Murray, M., Maeder, P., Fornari, E., Thiran, J., Clarke, S., Michel, C. & Meuli, R. (2006) Visuo-motor pathways in humans revealed by event-related fMRI. *Exp. Brain Res.*, **170**, 472–487.
- Martuzzi, R., Murray, M., Michel, C., Thiran, J.-P., Maeder, P., Clarke, S. & Meul, R. (2007) Multisensory interactions within human primary cortices revealed by BOLD dynamics. *Cereb. Cortex*, **17**, 1672–1679.
- Mesulam, M.-M. (1998) From sensation to cognition. *Brain*, **121**, 1013–1052.
- Miller, L. & D’Esposito, M. (2005) Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *J. Neurosci.*, **25**, 5884–5893.
- Mishra, J., Martinez, A., Sejnowski, T. & Hillyard, S. (2007) Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *J. Neurosci.*, **27**, 4120–4131.
- Molholm, S. & Foxe, J. (2005) Look ‘hear’, primary auditory cortex is active during lip-reading. *Neuroreport*, **16**, 123–124.

- Molholm, S., Ritter, W., Murray, M., Javitt, D., Schroeder, C. & Foxe, J. (2002) Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain Res. Cogn. Brain Res.*, **14**, 115–128.
- Molholm, S., Ritter, W., Javitt, D. & Foxe, J. (2004) Multisensory visual-auditory object recognition in humans: a high-density electrical mapping study. *Cereb. Cortex*, **14**, 452–465.
- de la Mothe, L., Blumell, S., Kajikawa, Y. & Hackett, T. (2006a) Cortical connections of the auditory cortex in marmoset monkeys: Core and medial belt regions. *J. Comp. Neurol.*, **496**, 27–71.
- de la Mothe, L., Blumell, S., Kajikawa, Y. & Hackett, T. (2006b) Thalamic connections of the auditory cortex in marmoset monkeys: Core and medial belt regions. *J. Comp. Neurol.*, **496**, 72–96.
- Murray, M., Molholm, S., Michel, C., Heslenfeld, D., Ritter, W., Javitt, D., Schroeder, C. & Foxe, J. (2005) Grabbing your ear: rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cereb. Cortex*, **15**, 963–974.
- Musacchia, G. & Schroeder, C. (2009) Neuronal mechanisms, response dynamics and perceptual functions of multisensory interactions in auditory cortex. *Hear. Res.*, **258**, 72–79.
- Musacchia, G., Sams, M., Nicol, T. & Kraus, N. (2006) Seeing speech affects acoustic information processing in the human brainstem. *Exp. Brain Res.*, **168**, 1–10.
- Parkkonen, L., Fujiki, N. & Mäkelä, J. (2009) Sources of auditory brainstem responses revisited: contribution by magnetoencephalography. *Hum. Brain Mapp.*, **30**, 1772–1782.
- Pekkola, J., Ojanen, V., Autti, T., Jaaskelainen, I., Mottonen, R., Tarkiainen, A. & Sams, M. (2005) Primary auditory cortex activation by visual speech: an fMRI study at 3 T. *Neuroreport*, **16**, 125–128.
- Poghosyan, V. & Ioannides, A. (2008) Attention modulates earliest responses in the primary auditory and visual cortices. *Neuron*, **58**, 802–813.
- Raij, T., Uutela, K. & Hari, R. (2000) Audiovisual integration of letters in the human brain. *Neuron*, **28**, 617–625.
- Raij, T., Karhu, J., Kičić, D., Lioumis, P., Julkunen, P., Lin, F., Ahveninen, J., Ilmoniemi, R., Mäkelä, J., Hämäläinen, M., Rosen, B. & Belliveau, J. (2008) Parallel input makes the brain run faster. *Neuroimage*, **40**, 1792–1797.
- Rockland, K. & Ojima, H. (2003) Multisensory convergence in calcarine visual areas in macaque monkey. *Int. J. Psychophysiol.*, **50**, 19–26.
- Rockland, K. & Van Hoesen, G. (1994) Direct temporal-occipital feedback connections to striate cortex (V1) in the macaque monkey. *Cereb. Cortex*, **4**, 300–313.
- Schoffelen, J. & Gross, J. (2009) Source connectivity analysis with MEG and EEG. *Hum. Brain Mapp.*, **30**, 1857–1865.
- Schroeder, C. & Foxe, J. (2002) The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Brain Res. Cogn. Brain Res.*, **14**, 187–198.
- Schroeder, C. & Foxe, J. (2005) Multisensory contributions to low-level, ‘unisensory’ processing. *Curr. Opin. Neurobiol.*, **15**, 454–458.
- Schroeder, C.E., Lindsley, R.W., Specht, C., Marcovici, A., Smiley, J.F. & Javitt, D.C. (2001) Somatosensory input to auditory association cortex in the macaque monkey. *J. Neurophysiol.*, **85**, 1322–1327.
- Schroeder, C., Smiley, J., Fu, K., McGinnis, T., O’Connell, M. & Hackett, T. (2003) Anatomical mechanisms and functional implications of multisensory convergence in early cortical processing. *Int. J. Psychophysiol.*, **50**, 5–17.
- Shams, L., Kamitani, Y. & Shimojo, S. (2002) Visual illusion induced by sound. *Brain Res. Cogn. Brain Res.*, **14**, 147–152.
- Shams, L., Iwaki, S., Chawla, A. & Bhattacharya, J. (2005) Early modulation of visual cortex by sound: an MEG study. *Neurosci. Lett.*, **378**, 76–81.
- Smiley, J. & Falchier, A. (2009) Multisensory connections of monkey auditory cerebral cortex. *Hear. Res.*, **258**, 37–46.
- Smiley, J., Hackett, T., Ulbert, I., Karmas, G., Lakatos, P., Javitt, D. & Schroeder, C. (2007) Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. *J. Comp. Neurol.*, **502**, 894–923.
- Sperdin, H., Cappe, C., Foxe, J. & Murray, M. (2009) Early, low-level auditory-somatosensory multisensory interactions impact reaction time speed. *Front. Integr. Neurosci.*, **3**, 2.
- Stein, B. & Meredith, M. (1993) *The Merging of the Senses*. MIT Press, Cambridge MA.
- Stein, B., Wallace, M., Stanford, T. & Jiang, W. (2002) Cortex governs multisensory integration in the midbrain. *Neuroscientist*, **8**, 306–314.
- Talsma, D. & Woldorff, M. (2005) Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. *J. Cogn. Neurosci.*, **17**, 1098–1114.
- Talsma, D., Doty, T. & Woldorff, M. (2007) Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? *Cereb. Cortex*, **17**, 679–690.
- Teder-Sälejärvi, W., McDonald, J., Di Russo, F. & Hillyard, S. (2002) An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Brain Res. Cogn. Brain Res.*, **14**, 106–114.
- Wang, Y., Celebrini, S., Trotter, Y. & Barone, P. (2008) Visuo-auditory interactions in the primary visual cortex of the behaving monkey: electrophysiological evidence. *BMC Neurosci.*, **9**, 79.
- Watkins, S., Shams, L., Tanaka, S., Haynes, J. & Rees, G. (2006) Sound alters activity in human V1 in association with illusory visual perception. *Neuroimage*, **31**, 1247–1256.