



OPEN ACCESS

EDITED BY

Benjamin A. Rowland,
Wake Forest University, United States

REVIEWED BY

Michael S. Beauchamp,
University of Pennsylvania, United States
Amir Borna,
Sandia National Laboratories (DOE),
United States

*CORRESPONDENCE

Tommi Raij
✉ raij@nmr.mgh.harvard.edu

PRESENT ADDRESS

Benjamin Letham,
Meta Inc., Data Science, Menlo Park, CA,
United States
Tapsya Nayak,
University of Texas Health Center at San
Antonio, TX, United States
Thomas Witzel,
Q Bio Inc., San Carlos, CA, United States

RECEIVED 03 May 2024

ACCEPTED 29 October 2024

PUBLISHED 14 November 2024

CITATION

Raij T, Lin F-H, Letham B, Lankinen K,
Nayak T, Witzel T, Hämäläinen M and
Ahveninen J (2024) Onset timing of letter
processing in auditory and visual sensory
cortices.
Front. Integr. Neurosci. 18:1427149.
doi: 10.3389/fnint.2024.1427149

COPYRIGHT

© 2024 Raij, Lin, Letham, Lankinen, Nayak,
Witzel, Hämäläinen and Ahveninen. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Onset timing of letter processing in auditory and visual sensory cortices

Tommi Raij^{1*}, Fa-Hsuan Lin^{2,3,4}, Benjamin Letham^{1†},
Kaisu Lankinen¹, Tapsya Nayak^{1†}, Thomas Witzel^{1,5†},
Matti Hämäläinen⁶ and Jyrki Ahveninen¹

¹MGH/MIT/HMS Athinoula A. Martinos Center for Biomedical Imaging, Charlestown, MA, United States, ²Physical Sciences Platform, Sunnybrook Research Institute, Toronto, ON, Canada, ³Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada, ⁴Institute of Medical Science, University of Toronto, Toronto, ON, Canada, ⁵Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA, United States, ⁶Department of Neuroscience and Biomedical Engineering, Aalto University School of Science, Espoo, Finland

Here, we report onset latencies for multisensory processing of letters in the primary auditory and visual sensory cortices. Healthy adults were presented with 300-ms visual and/or auditory letters (uppercase Roman alphabet and the corresponding auditory letter names in English). Magnetoencephalography (MEG) evoked response generators were extracted from the auditory and visual sensory cortices for both within-modality and cross-sensory activations; these locations were mainly consistent with functional magnetic resonance imaging (fMRI) results in the same subjects. In the primary auditory cortices (Heschl's gyri) activity to auditory stimuli commenced at 25 ms and to visual stimuli at 65 ms (median values). In the primary visual cortex (Calcarine fissure) the activations started at 48 ms to visual and at 62 ms to auditory stimuli. This timing pattern suggests 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 17–37 ms. Audiovisual interactions for letters started at 125 ms in the auditory and at 133 ms in the visual cortex (60–71 ms after inputs from both modalities converged). Multivariate pattern analysis suggested similar latency differences between the sensory cortices. Combined with our earlier findings for simpler stimuli (noise bursts and checkerboards), these results suggest that primary sensory cortices participate in early cross-modal and interaction processes similarly for different stimulus materials, but previously learned audiovisual associations and stimulus complexity may delay the start of the audiovisual interaction stage.

KEYWORDS

audiovisual interaction, cross-modal, language, MEG, multisensory

1 Introduction

Letters of the alphabet are the basic building blocks of written phonetic language. The associations between auditory and visual letters are arbitrary, language-dependent, and based on extensive learning. For these reasons, brain representations of letters have been used in studies of multisensory processing, learning, distributed representations of supramodal concepts, and dyslexia (Raij, 1999; Raij et al., 2000; van Atteveldt et al., 2004; Herdman et al., 2006; Blau et al., 2010; Andres et al., 2011; Blomert, 2011; Froyen et al., 2011).

While the traditional view was that primary sensory areas show strict sensory fidelity, i.e., they can only be activated by stimuli in the appropriate sensory modality, it is now widely accepted that even low-order sensory areas may show cross-sensory (i.e., cross-modal)

activations and multisensory interactions starting as early as about 40–50 ms after stimulus onset, in both nonhuman primates (Schroeder et al., 2001; Schroeder and Foxe, 2002) and humans (Giard and Peronnet, 1999; Foxe et al., 2000; Molholm et al., 2002; Teder-Sälejärvi et al., 2002; Molholm et al., 2004; Murray et al., 2005; Talsma et al., 2007; Cappe et al., 2010; Raij et al., 2010; Beer et al., 2013). Supporting evidence comes from fMRI results showing cross-sensory activations in or very close to primary sensory areas (Pekkola et al., 2005; Martuzzi et al., 2006; Raij et al., 2010). These findings are consistent with functional and structural MRI connectivity analyses indicating direct connections between the primary auditory and visual cortices (Eckert et al., 2008; Beer et al., 2011; Beer et al., 2013). This suggests that low-order sensory areas may contribute to multisensory integration starting from very early processing stages (Schroeder et al., 2003; Foxe and Schroeder, 2005; Macaluso and Driver, 2005; Molholm and Foxe, 2005; Schroeder and Foxe, 2005; Ghazanfar and Schroeder, 2006; Macaluso, 2006; Kayser and Logothetis, 2007; Musacchia and Schroeder, 2009; Raij et al., 2010). However, the types of stimuli utilized in these studies have been limited, and it is unclear if audiovisual learning influences these processes. Consequently, the functional roles of such early multisensory activations remain elusive.

Determining onset latencies of stimulus-evoked responses is a robust way of detecting the order and spread of activations across different brain areas. We have previously reported onset latencies for audiovisual processing of noise bursts and checkerboard stimuli (Raij et al., 2010). However, such simple stimuli, while strongly activating sensory systems, do not have learned audiovisual associations, which may influence multisensory interactions (Raij et al., 2000; van Atteveldt et al., 2004). Letters are physically relatively simple and therefore allow meaningful comparisons with studies utilizing simpler stimuli; yet, they have strong audiovisual associations formed through extensive learning. Therefore, we here examine onset latencies for letters of the alphabet. To improve comparability between studies, the subjects, experimental design, and recordings are identical to our earlier experiment with simpler stimuli (Raij et al., 2010), together with the present data forming a multimodal MRI/fMRI/MEG dataset on multisensory processing. The present data were recorded in the same session as in Raij et al. (2010) and were previously used in a different study (Lankinen et al., 2024).

2 Materials and methods

2.1 Subjects, stimuli, and tasks

The protocol was approved by the Massachusetts General Hospital institutional review board, and subjects gave their written informed consent prior to participation. The visual stimuli were individual uppercase letters of the Roman alphabet (visual angle $3.5^\circ \times 3.5^\circ$, contrast 100%, foveal presentation) and auditory stimuli the corresponding spoken English letter names. Sixteen different letters were used (ADEFIKLMNORSTVXY). The auditory stimuli were recorded from a female speaker in an echoless chamber at the Department of Cognitive and Neural Systems at Boston University. The duration of all stimuli was 300 ms. Auditory stimulus onset was defined as the onset of the first audible component for each letter; as expected, the envelope profiles following onsets varied across auditory letters. The letters were presented as auditory only (A), visual only (V),

or audiovisual combination (AV, simultaneous auditory and visual, always congruent), in a rapid event-related fMRI-type design with pseudorandom stimulus order and interstimulus interval (ISI). A/V/AV stimuli were equiprobable. Subjects were 8 healthy right-handed humans (6 females, age 22–30). The task was to respond to rare (10%) auditory (sound [kei]), visual (letter K), or audiovisual ([kei]/K) target stimuli (A Target / V Target / AV Target) with the right index finger as quickly as possible while reaction time (RT) was measured. All subjects were recorded with three stimulus sequences having different interstimulus intervals (ISIs) with mean ISIs at 1.5/3.1/6.1 s, across which the MEG onset latencies were practically identical and were therefore averaged within subjects. Inside each sequence the ISI was jittered at 1.15 s (1 TR of the fMRI acquisition) to improve fMRI analysis power (Dale, 1999; Burock and Dale, 2000). Identical stimuli and tasks were used in MEG and fMRI. Visual stimuli were projected with a video projector on a translucent screen. The auditory stimuli were presented with MEG-compatible headphones or through MRI-compatible headphones (MR Confon GmbH, Magdeburg, Germany). Auditory stimuli were adjusted to be as loud as the subject could comfortably listen to (in MEG about 65 dB SPL; in fMRI clearly above the scanner acoustical noise). Stimuli were presented with a PC running Presentation 9.20 (Neurobehavioral Systems Inc., Albany, CA, USA.). In fMRI stimuli were synchronized with triggers from the fMRI scanner. Timing of the stimuli with respect to trigger signals was confirmed with a digital oscilloscope. To maximize comparability, all parameters (except the stimuli) were identical to those used in our previous study utilizing simpler stimuli (Raij et al., 2010).

2.2 Structural MRI recordings and analysis

Structural T1-weighted MRIs were acquired with a 1.5 T 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 (Fischl et al., 2002; Fischl et al., 2004). Individual brains were spatially co-registered by morphing them into the FreeSurfer average brain via a spherical surface (Fischl et al., 1999).

2.3 fMRI recordings and analysis

Brain activity was measured using a 3.0 T Siemens Trio scanner with a Siemens head coil and an echo planar imaging (EPI) blood oxygenation level dependent (BOLD) sequence (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). fMRI data were analyzed with FreeSurfer. During preprocessing, data were motion corrected (Cox and 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 1,000. The first three images of each run were excluded 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 and 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 / AV Target) with a time window of 2.3 s pre-stimulus to 16.1 s post-stimulus. 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. To enhance comparability, these parameters were identical to those used in our previous study (Raij et al., 2010).

2.4 MEG recordings and alignment with MRI

The MEG equipment, recordings, and analyses were the same as those reported previously (Raij et al., 2010). Whole-head 306-channel MEG (VectorView, Elekta-Neuromag, Finland) was recorded in a magnetically shielded room (Cohen et al., 2002; Hämäläinen and 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 band-pass filtered to 0.03–200 Hz prior to sampling at 600 Hz. Prior to the MEG recordings, the locations of four small head position indicator (HPI) 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). 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.

2.5 MEG evoked response analysis (sensor space)

The MEG responses were averaged offline separately for each trial type (A/V/AV/A Target/V Target/AV Target), time locked to the stimulus onsets, from 250 ms pre-stimulus to 1,150 ms post-stimulus. A total of 375 individual trials per category were recorded for the non-target conditions (100 epochs for the long, 125 for the intermediate, and 150 for the short ISI run). Trials exceeding 150 μ V or 3,000 fT/cm at any EOG or MEG channel, respectively, were automatically discarded. The averaged signals were digitally low-pass filtered at 40 Hz and amplitudes measured with respect to a 200-ms pre-stimulus baseline. Onset timings were analyzed separately for the responses to A, V, and AV stimuli. Specifically, for the sensor space analysis, we estimated onsets from the amplitudes $\sqrt{(b_x^2 + b_y^2)}$ computed from the signals b_x and b_y of the two planar gradiometers at each sensor location (Supplementary material for details). From each subject, 3 sensor locations were selected showing maximal responses over the primary auditory cortices (1 location in each hemisphere) and the primary visual cortex (1 location at posterior midline). Onset latencies were picked at the first time point that exceeded three standard deviations (3SDs) above noise level estimated from the 200-ms pre-stimulus baseline. In addition, we required that the response must not start earlier than 15 ms (based on finite conduction delays in sensory pathways) and the response has to stay above the noise level for at least 20 ms (to protect against brief noise spikes). The onset latency analyses were done separately for the grand average time course (averaged across all accepted conditions and subjects to improve the signal-to-noise ratio SNR) and for the individual level responses (to allow computation of

variability across subjects but with lower SNR). Data from one subject were too noisy for accurate onset latency determination and were therefore discarded; the same subject was discarded in the previous publication utilizing simpler stimuli. Additionally, in two subjects, one run (out of the total of three runs with different ISIs) was contaminated by eye blinks and therefore discarded; in these subjects only the data from the remaining two runs were used. After combining the runs with different ISIs, each subject's averaged response consisted of about 300 trials (for details see Supplementary material).

Finally, to estimate AV interactions, we calculated the AV interaction responses $[AV - (A + V)]$ from their constituent A, V, and AV evoked responses. Such responses have been used widely to study multisensory interactions since their introduction (Morrell, 1968). However, addition and subtraction of responses decreases signal-to-noise ratios (for interaction responses, theoretically by $\sqrt{3}$ times over their constituent A, V and AV responses). Therefore, the AV interaction responses were filtered more strictly (low-pass filtering at 20 Hz with 3 dB roll-off) and were excluded from sensor space analysis for onset latencies; instead, the sensor space AV interaction responses were subjected to source analysis (see below) followed by extraction of their onsets.

2.6 MEG source analysis and onset latencies (source space)

As in our earlier study (Raij et al., 2010), minimum-norm estimates (MNEs) (Hämäläinen and Ilmoniemi, 1984; Hämäläinen and Ilmoniemi, 1994) were computed from combined anatomical MRI and MEG data (Dale and Sereno, 1993; Liu et al., 1998; Dale et al., 2000) using the MNE software (Gramfort et al., 2014). The noise normalized MNE (dSPM) values were then calculated to reduce the point-spread function and to allow displaying the activations as an *F*-statistic. 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. The dSPM time courses were calculated separately for A, V, AV, and AV interaction responses and extracted from anatomically pre-determined (Desikan et al., 2006) locations of A1 and V1. Next, their onset latencies were measured as described above for sensor signals using the 3SD and latency minimum criteria. These values were extracted separately from grand average time courses (averaged across subjects) and from the individual level data. Finally, to extract the onset latencies from the source-space signals with high SNR while also being able to estimate their variances, we additionally used bootstrapping (for details see Supplementary material).

2.7 MVPA analysis of MEG source space data

We also examined early cross-sensory influences using multivariate pattern analysis (MVPA). MVPA decoding analysis was performed by two-class support vector machine (SVM) classifier with linear kernel and cost equal to one ($C=1$) implemented in libsvm (Chang and Lin, 2011) and provided in the CoSMoMVPA package¹ (Oosterhof et al., 2016) in MATLAB

¹ <http://www.cosmormvpa.org/>

(Natick, MA, USA). The decoding was performed separately for each subject, with bilateral A1:s and V1:s as the ROIs. The individual-trial MEG source estimates were averaged within each hemisphere-specific ROI before MVPAs. For this analysis, only the two runs with longer ISIs were selected, because the recordings with the shortest ISI would have had crosstalk between trials. The classification was conducted using a temporal searchlight analysis with a sliding window width of 50 ms with the window moved at 1.7 ms steps (the sampling rate). For contrasts that required a “noise” sample, the noise data were drawn from the time segments between stimuli. In each of our 100 randomized cross-validation folds, the model was trained in 80% of the trials and tested in the remaining 20% of trials. The decoding accuracy was averaged across the 100 folds. We tested the statistical significance for each subject using a *t*-test with a chance level of 0.5. Finally, the *p*-values were corrected for multiple comparisons using the false discovery rate (FDR) procedure (Benjamini and Hochberg, 1995).

3 Results

3.1 Behavioral results

The average hit rate to target stimuli was 97% across all experimental conditions. During MEG, the RTs were faster for AV (median 487 ms, mean \pm SD 492 \pm 65 ms) than for A (median 576 ms, mean \pm SD 579 \pm 99 ms) and V (median 518 ms, mean \pm SD 530 \pm 73 ms) stimuli with outliers excluded according to the Median Absolute Deviation (MAD) statistics criterion. In fMRI the difference was slightly smaller (for AV median 599 ms, mean 603 \pm 80 ms, for A median 700 ms, mean 722 \pm 128 ms, and for V stimuli median 615 ms, mean 623 \pm 86 ms). Similar as in our earlier study (Rajj et al., 2010), the longer RTs in fMRI may be due to slower response pads and the MR environment. Since the A, V, and AV stimuli were in random order and stimulus timing was pseudorandom, attention related or anticipatory differences could not have influenced the onset latency differences across conditions.

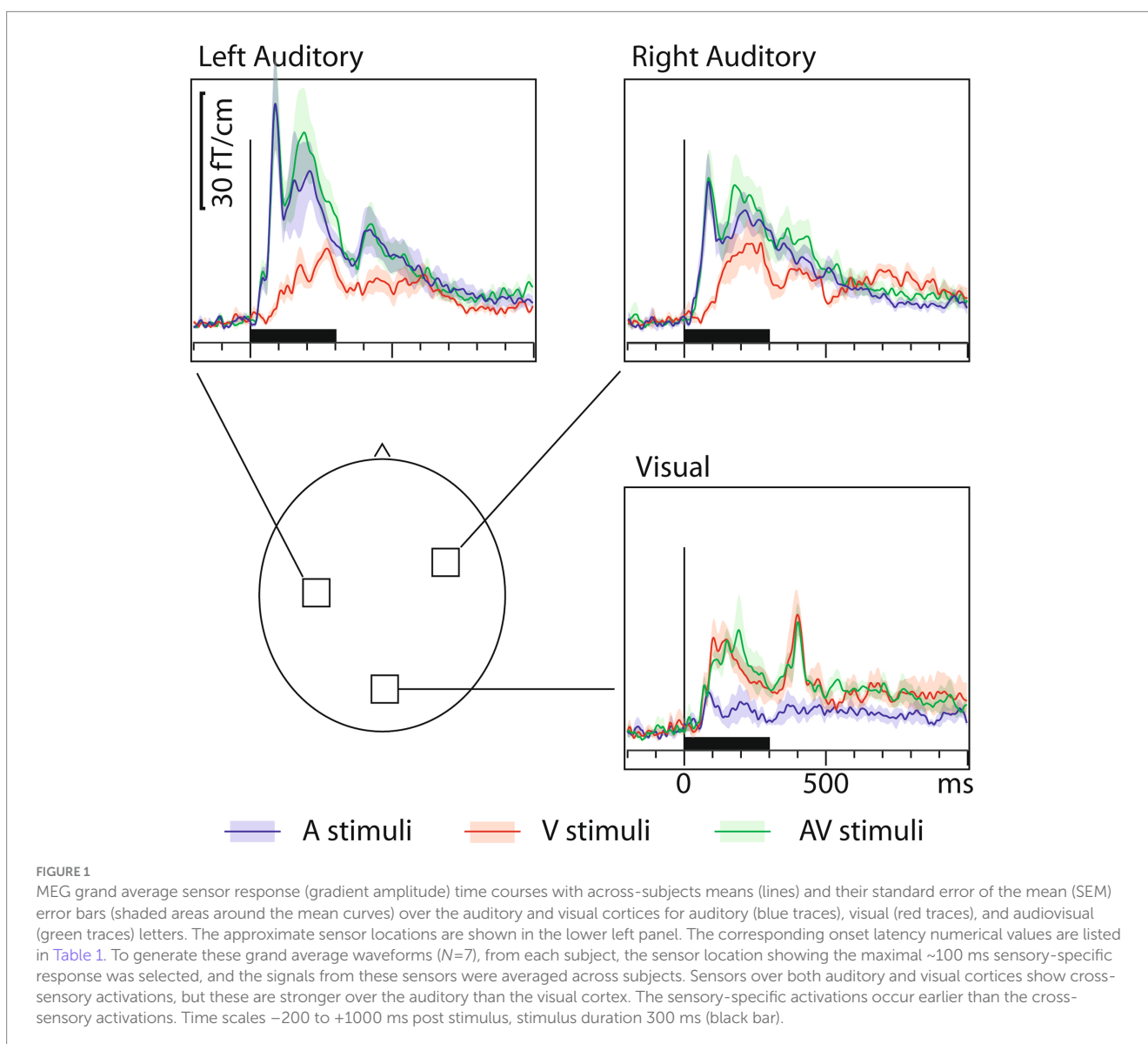


TABLE 1 Sensor onset latencies (ms) over the auditory and visual cortices.

Sensor latencies (ms) Grand average responses	Auditory cortex		Visual cortex
	Left	Right	
Auditory stimuli	27	25	63
Visual stimuli	80	70	60
AV stimuli	27	30	55

The values were picked from the grand average time courses shown in Figure 1; for individual subject values, see Table 2.

TABLE 2 Across-subjects sensor onset latencies (ms) over the auditory and visual cortices.

Sensor latencies (ms) Across-subjects values	Auditory cortex Mean \pm SD (median)	Visual cortex Mean \pm SD (median)
Auditory stimuli	38 \pm 17 (32)	76 \pm 39 (65)
Visual stimuli	91 \pm 11 (93)	63 \pm 10 (65)
AV stimuli	39 \pm 15 (33)	66 \pm 31 (57)

These values were picked from the individual level responses.

3.2 MEG onset latencies: sensor data

Figure 1 shows that, in addition to the expected sensory-specific activations, cross-sensory effects were observed: visual stimuli strongly activated temporal cortices and auditory stimuli (albeit more weakly) the midline occipital cortex. Table 1 lists the corresponding onset latencies (the time when the grand average response first exceeded 3SD above the noise level). The sensory-specific activations started 34 ms earlier over the auditory than visual cortex. The cross-sensory activations started after the sensory-specific responses, by 3 ms over visual cortex and by 49 ms over auditory cortex. Table 2 lists the across-subjects onset latencies; the left and right auditory cortices showed similar timings and were thus averaged. The individual subjects' sensor-level responses were relatively noisy and thus did not correspond well to the grand average results; hence, no statistical comparisons were done for the sensor data (see dSPM data below for statistical tests across individuals).

3.3 MEG dSPM source analysis

Figure 2 shows the MEG localization results at selected time points after the onset of activity. Similar as for sensor space analysis, in addition to sensory-specific activations, cross-sensory activations were observed: visual stimuli strongly activated large areas of temporal cortex including the supratemporal auditory cortex, and auditory 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 the primary sensory areas.

3.4 MEG dSPM source-specific onset latencies

Figure 3 shows MEG dSPM time courses from Heschl's gyri (HG/A1) and calcarine fissure (V1) for auditory, visual, and AV stimuli. The left and right calcarine fissure activations, due to their close anatomical proximity and similar timings, were averaged. As expected, the time courses are similar to the sensor amplitudes in Figure 1. However, in the presence of multiple sources, dSPM time courses more accurately allow extraction of activity from a specific area than sensors that collect activity from a rather large area. Table 3 lists the corresponding onset latencies measured from the grand average dSPM responses using bootstrapping. Since onsets were very similar across hemispheres, the responses were averaged across the left and right hemisphere. Sensory-specific activations started 23 ms (median) earlier in A1 than in V1. In V1, sensory-specific activations started 14 ms before the cross-specific responses, whereas in A1 cross-sensory activations started 40 ms after the sensory-specific responses. Cross-modal conduction delays (spread from one sensory cortex to another) were 37 ms for auditory and 17 ms for visual stimuli. As expected, onsets of responses to AV stimuli closely followed onsets to the unimodal stimulus that first reached the sensory cortex.

Figure 4 shows dSPM time courses calculated from the audiovisual interaction responses. As expected, these were weaker than the constituent A/V/AV responses. In the auditory cortex the interactions started 60 ms after and in the visual cortex 71 ms after inputs from both sensory modalities converged.

Table 4 lists dSPM onset latencies (mean \pm SD and median) across the individual subjects (without bootstrapping). On the one hand, since these values were picked from the individual subject responses where SNR is lower than in group-level averaged responses, the onsets are somewhat longer than in Table 3. On the other hand, the individual level values allow for straightforward statistical testing. Sensory-specific auditory evoked responses in A1 started 28 ms earlier than visual evoked responses in V1 (Wilcoxon signed rank test for medians ($n=7$), $p=0.018$). Cross-sensory activations in A1 occurred 52 ms later than sensory-specific activations, which was statistically significant ($p=0.028$). In V1, cross-sensory activations occurred 10 ms later than sensory-specific activations, but this difference was not quite significant ($p=0.063$). Difference between cross-modal conduction delays (from one sensory cortex to another) for auditory and visual stimuli was non-significant ($p=0.128$).

3.5 MVPA results

Figure 5 shows the MVPA decoding results for relevant contrasts in A1 and V1. The results suggest similar processing stage and inter-regional latency differences between A1 and V1 as the ERP onsets.

3.6 fMRI activations

MEG source analysis can be complicated by the electromagnetic inverse problem and volume conduction. We therefore recorded fMRI activations using the same subjects and stimuli. Figure 6 shows the grand average fMRI results, with the BOLD time courses in the lower panels. Calcarine cortex was activated by both visual and auditory

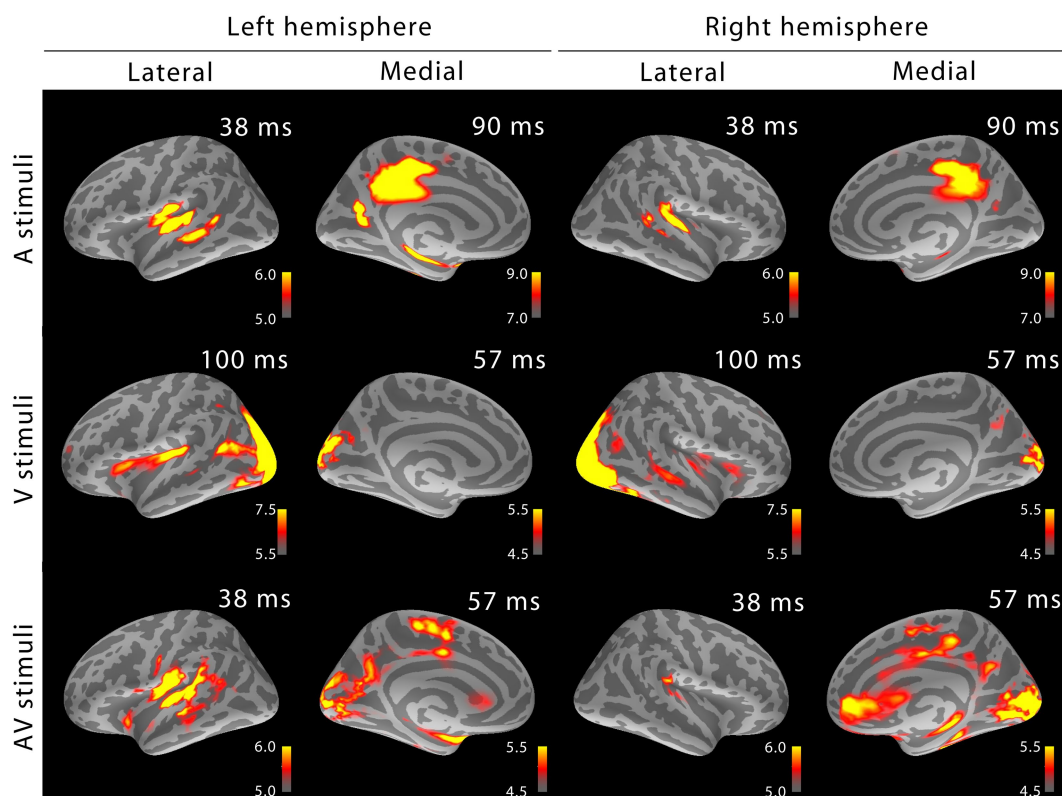


FIGURE 2

MEG snapshots (dSPM \sqrt{F} -statistics) at early activation latencies. Both sensory-specific and cross-sensory activations are seen at primary sensory areas (the right 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), other activations occupy somewhat different locations than the sensory-specific activations. However, the spatial resolution of MEG is somewhat limited – hence exact comparisons are discouraged. Outside sensory areas, the cross-sensory activations to visual stimuli show additional bilateral activations in superior temporal sulci (STS, more posterior in the left hemisphere) and left Broca's area. Grand average data ($N=7$).

letters, and in fact more strongly by the latter. Heschl's gyri showed strong activity for auditory stimuli, but, in contrast to the MEG results, for visual stimuli only the right hemisphere showed a small deflection at the typical BOLD signal peak latency. At closer inspection the right A1 voxels that were activated by visual letters were concentrated in the posteromedial part of Heschl's gyrus, spatially diluting this effect when averaged across the entire ROI.

3.7 Comparisons for activations between letters and simpler stimuli

We have previously reported onset latencies for simpler stimuli (checkerboards and noise bursts) in the same subjects using identical paradigms and recordings as in the present study (Raj et al., 2010). The values for sensory-specific and cross-sensory activations were similar regardless of stimulus type, as can be seen comparing the values between (Raj et al., 2010) and the present study; differences were non-significant, with the exception of V1 responses to visual stimuli starting later (by 13 ms for median, 8 ms for mean) for letters than for checkerboard stimuli (Wilcoxon Signed Rank Test for medians $p=0.034$). This difference can probably be attributed to that letters cover a smaller proportion of the visual field than

checkerboards, therefore activating fewer V1 neurons and weakening the population response.

The largest latency difference was that the AV interactions started later for letters (present study) than for simpler stimuli (Raj et al., 2010). In A1 this difference was 40 ms and in V1 56 ms (bootstrapped medians). These differences were statistically significant ($p < 0.001$ separately for both A1 and V1 for the bootstrapped medians; for details of the employed statistical test see Supplementary material).

Some source distribution/amplitude differences were also evident when comparing Figures 2 between the two studies. For sensory-specific activations, in Heschl's gyri the auditory letters evoked stronger responses than noise bursts, whereas in the Calcarine fissure, checkerboards resulted in stronger activation than visual letters. Source distributions outside primary sensory areas at early MEG latencies (Figure 2) were also somewhat different between basic stimuli and letters. Specifically, visual letters evoked a more left-lateralized response in Broca's area and stronger responses in left posterior superior temporal sulcus (STS) than checkerboards.

In fMRI, results were also similar across the two stimulus types, but some differences were found in the activation strengths. In V1, visual evoked activations were clearly stronger for checkerboards than for letters. This may reflect that the different letters activate smaller and retinotopically largely non-overlapping parts of the visual field

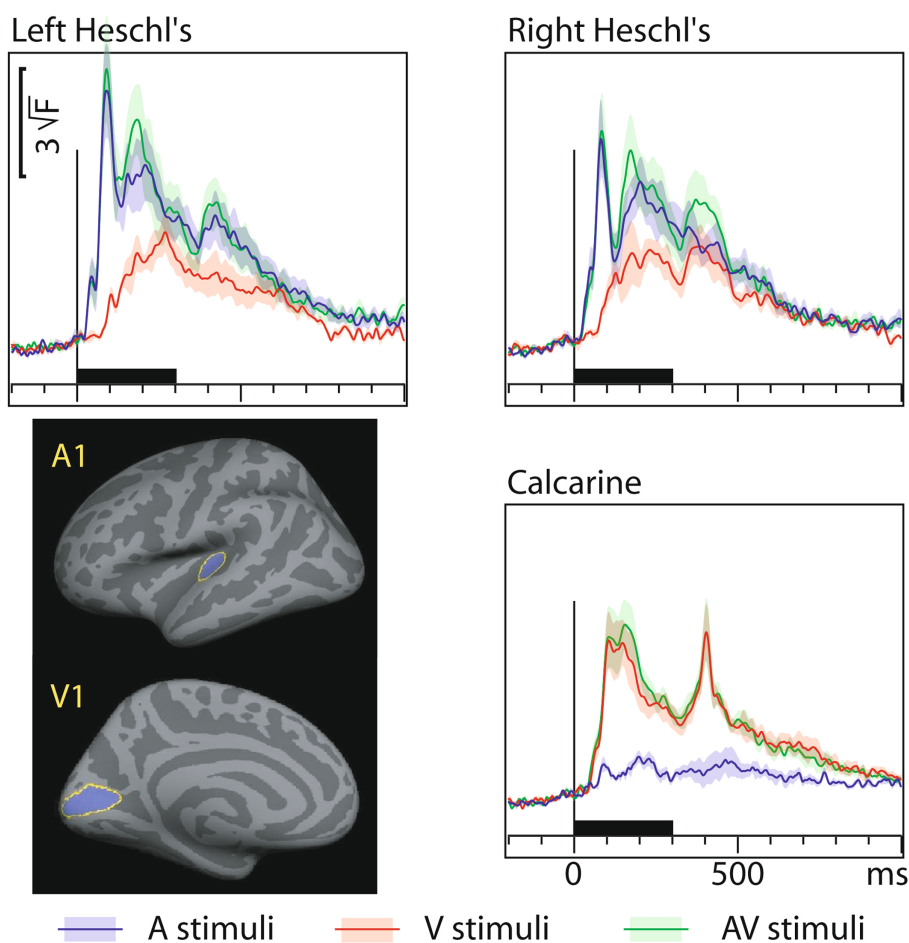


FIGURE 3
MEG grand average source-specific dSPM time courses and their across-subjects SEM error bars for Heschl's gyri (A1) and calcarine fissure (V1) to auditory, visual, and audiovisual stimuli. The source areas, shown for the left hemisphere in the lower left panel, were based on an anatomical parcellation (Desikan et al., 2006); left and right calcarine sources were averaged. The corresponding onset latency numerical values are listed in Table 3. Both sensory-specific and cross-sensory activations are observed. The sensory-specific activations occur earlier than the cross-sensory activations. Time scales -200 to +1000 ms post stimulus, stimulus duration 300 ms (black bar). Grand average data (N=7) showing means (lines) and SEM error bars (shaded areas around the mean time courses).

TABLE 3 Across-subjects onset latency means ± SDs (Medians) in source space.

Source latencies (ms) Across-subjects values using bootstrapping	Heschl's gyrus Mean ± SD (Median)	Calcarine cortex Mean ± SD (Median)
Auditory stimuli	25 ± 2 (25)	55 ± 18 (62)
Visual stimuli	58 ± 17 (65)	48 ± 6 (48)
AV stimuli	25 ± 2 (25)	46 ± 4 (45)
AV Interaction (20 Hz LPF)	131 ± 27 (125)	126 ± 56 (133)

In this analysis the evoked responses from each hemisphere and subject were bootstrapped. In addition, AV interaction responses were low-pass filtered at 20 Hz to mitigate SNR problems caused by the $[AV - (A + V)]$ operation. For the corresponding values for A/V/AV stimuli extracted from the individual subject values without bootstrapping, see Table 4.

than a checkerboard. Further, V1 was activated more strongly by auditory letters than by noise bursts, which may reflect that auditory letters possess learned audiovisual associations. Both of these factors

contributed to that in V1, auditory letters caused stronger activations than visual letters (i.e., cross-sensory activations were stronger than sensory-specific responses), which was not the case for the simpler stimuli. In A1, the responses to auditory letters and noise bursts were quite similar in amplitude. Further, in A1 the visual letters evoked weak BOLD activations only in the right hemisphere; for checkerboards similar weak activations were observed bilaterally.

4 Discussion

Early cross-sensory activations and audiovisual interactions were found in both A1 and V1. In A1 the delay from sensory-specific to cross-sensory activity was 40 ms, whereas in V1 the delay was only 14 ms (see also (Brang et al., 2015) intracranial data showing about simultaneous activation of human V1 by auditory and visual stimuli). This asymmetrical timing pattern, where sensory-specific activations start earlier in A1 (25 ms) than in V1 (48 ms) cortex, is consistent with that the cross-sensory activations originate in the sensory cortex of the opposite

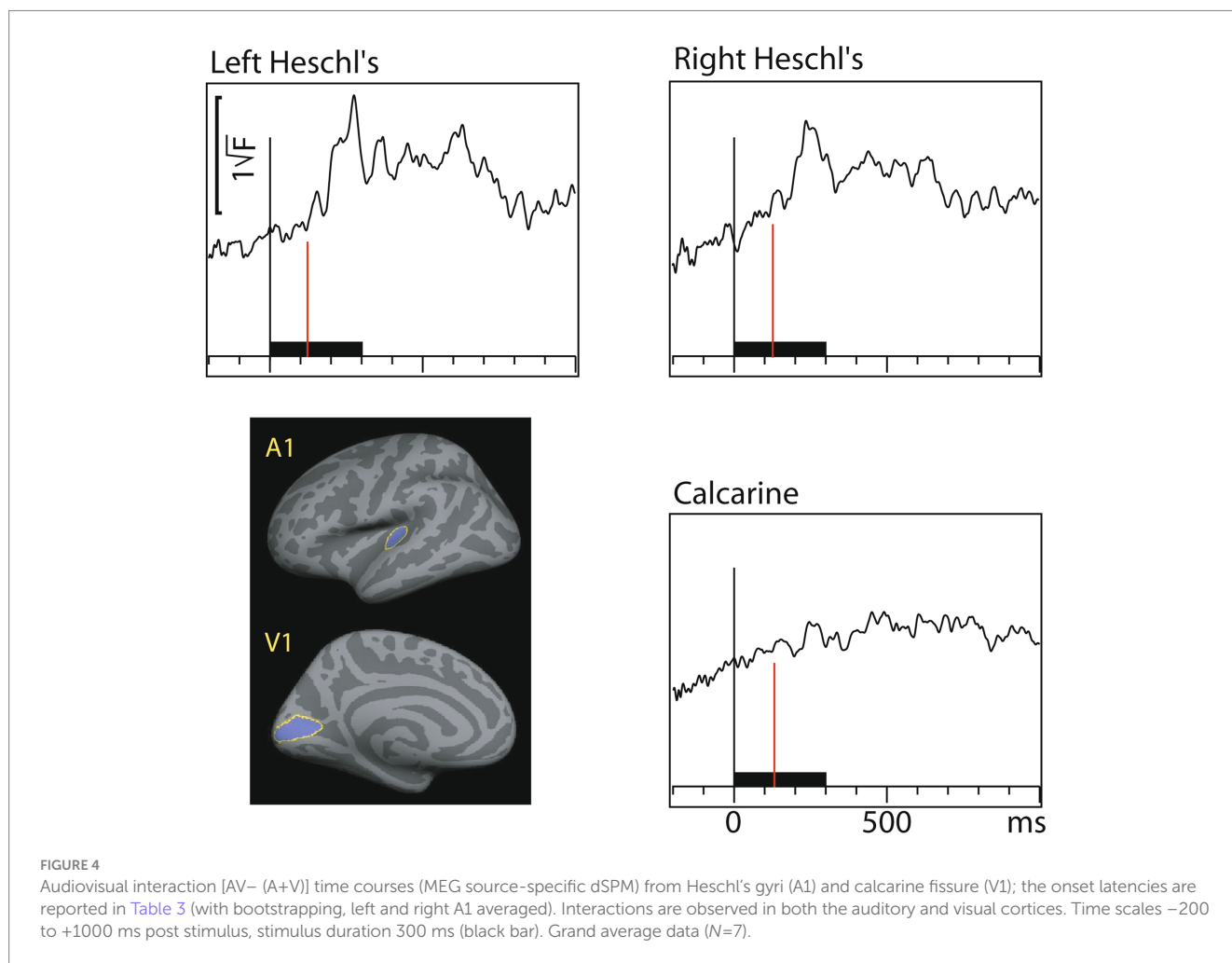


TABLE 4 Across-subjects source-specific onset latencies (ms).

dSPM latencies (ms) Across-subjects values	Heschl's gyrus Mean \pm SD (Median)	Calcarine cortex Mean \pm SD (Median)
Auditory stimuli	32 \pm 13 (30)	94 \pm 52 (68)
Visual stimuli	82 \pm 16 (82)	56 \pm 9 (58)
AV stimuli	32 \pm 12 (28)	51 \pm 13 (55)

These values were picked from the individual level responses.

stimulus modality, with about 17–37 ms conduction delay between A1 and V1. Audiovisual interactions were observed only after both sensory-specific and cross-sensory inputs converged on the sensory cortex.

The present results expand our earlier results obtained using simpler stimuli (Rajj et al., 2010; Ahveninen et al., 2024). Despite clearly different stimuli, the sensory-specific and cross-sensory onset latencies were quite consistent between letters and simpler stimuli. The cross-modal conduction delays were also similar. It remains plausible that the earliest cross-sensory activations may utilize the A1 \rightarrow V1 and V1 \rightarrow V2 \rightarrow A1 pathways. However, other viable options remain, including other cortico-cortical connections between the auditory and visual cortices, connections through

subcortical relays, or through an association cortical area such as STP/STS (Foxe and Simpson, 2002; Schroeder and Foxe, 2002; Schroeder et al., 2003; Liang et al., 2013; Henschke et al., 2015; Lohse et al., 2021). Perhaps in support of the last option, in the present data, left posterior STS was strongly activated at the same time as the cross-sensory auditory cortex activation occurred. However, in the right hemisphere only weak STS activity was observed. This left STS hemispheric lateralization is reverse to what was observed for simpler stimuli; it may be that different stimulus materials utilize partially different pathways, with language stimuli lateralized to the left. Further studies are needed to characterize the timing of network nodes supporting multisensory processing outside the primary sensory areas.

In accord with our previous results (Rajj et al., 2010), audiovisual interactions started only after the sensory-specific and cross-sensory inputs converged at the sensory cortex (at 80 ms in the auditory cortex and 50 ms in the visual cortex, i.e., in the auditory cortex 40 ms and in the visual cortex 77 ms after convergence). However, these interactions started later for letters (present study) than for simpler stimuli (Rajj et al., 2010). Whether this difference reflects physical stimulus properties (e.g., the amplitude envelopes of spoken letter names were more varied than noise bursts) or that previously learned audiovisual associations

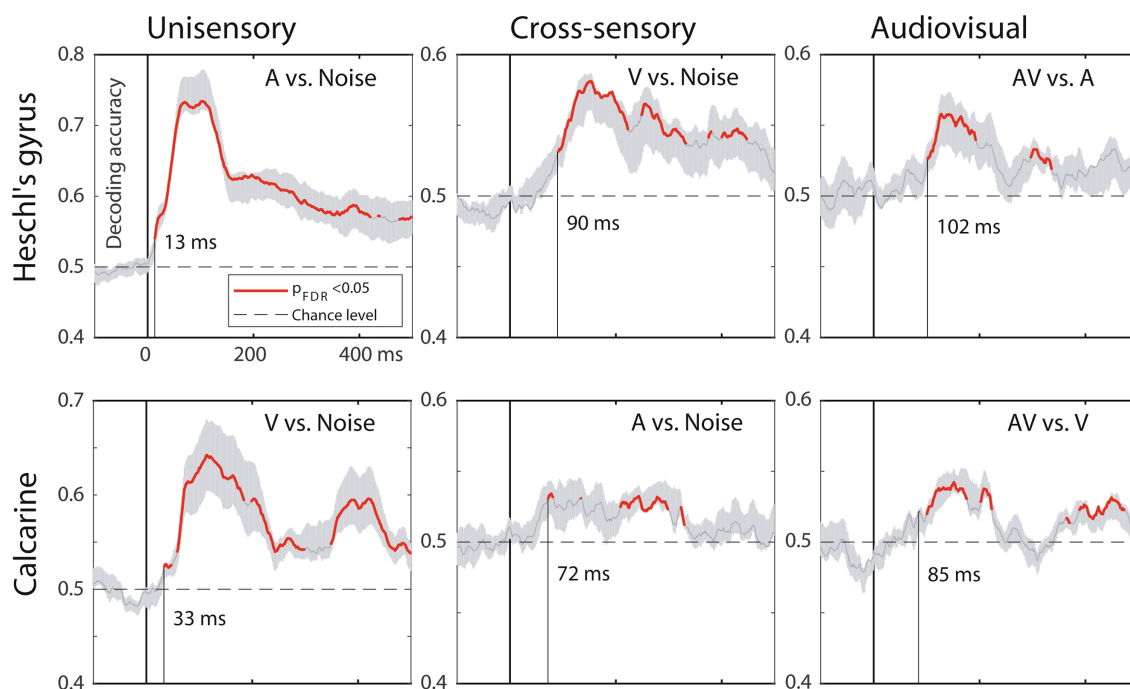


FIGURE 5

MVPA results, MEG source-space analysis in A1 (top) and V1 (bottom). The traces depict decoding accuracy over time (red = accuracy exceeds the statistical significance threshold; 0 ms = stimulus onset; post-stimulus vertical line = time when the decoding accuracy first becomes significant; grey shade = SEM across subjects). Decoding for unimodal stimuli starts in sensory-specific cortices (left column) and emerges faster in A1 than in V1. Cross-sensory decoding for unimodal stimuli (middle column) occurs later, followed by audiovisual interactions for bimodal stimuli (right column). Time scale -100 to $+500$ ms post stimulus, stimulus duration 300 ms, analysis time window length 50 ms, time window sliding step 1.7 ms.

prolong the onset of interactions warrants further investigation. Yet, prior electroencephalography (EEG) evidence shows that audiovisual interactions in humans can occur earlier, starting already at about 40 ms, over posterior areas (Giard and Peronnet, 1999; Molholm et al., 2002; Teder-Sälejärvi et al., 2002; Molholm et al., 2004; Cappe et al., 2010). A possible explanation is that the earliest interactions in EEG may be generated in subcortical structures (Rajj et al., 2010); EEG is more sensitive to signals from deep structures than MEG (Goldenholz et al., 2009).

The MVPA results mainly agreed with the evoked response onset latencies in terms of relative timing differences between A1 and V1 and the order of processing stages. However, it is worth noting that these MVPA latencies refer to the centroid of a sliding time window, which was utilized to classify different sensory conditions based on a dynamic pattern of brain activity. Thus, while MVPA decoding provided strong evidence of cross-modal influences in auditory and visual cortices, this analysis technique did not provide direct information on the actual neurophysiological onset latencies. Evoked response onset latencies remain better suited for this purpose.

The fMRI results largely agreed with the MEG localization results, but some discrepancies were observed. In V1, fMRI detected strong cross-sensory activations, which was in accord with MEG results. Moreover, in V1 the cross-sensory activations were stronger than the sensory-specific activations, probably reflecting that the visual letters only activated a small part of the visual field [compare these to the cross-sensory activations to checkerboards in Rajj et al., 2010]. However, in A1, where MEG showed strong cross-sensory

activations, fMRI showed only weak right hemisphere (for letters) or bilateral (for checkerboards) cross-sensory responses. Summarizing the fMRI findings from these two studies, it seems likely that A1 may be activated by visual stimuli. This is in agreement with an fMRI study reporting more robust A1 activations for checkerboards (Martuzzi et al., 2007), another fMRI study reporting that responses to auditory letters are modulated by simultaneously presented visual letters (van Atteveldt et al., 2004), and with our MEG findings. One explanation why the fMRI cross-sensory A1 activations in our two studies are weak may be that the acoustical scanner noise, accentuated by our rapid scanning parameters (short TR), dampened auditory cortex responses due to neuronal adaptation (in contrast, the MEG scanner is silent). Future fMRI studies using acoustically more quiet continuous EPI sequences or sparse sampling may offer further insight (Hall et al., 1999; Hennel et al., 1999; Yang et al., 2000; Schwarzbauer et al., 2006; Gaab et al., 2007a,b; Schmitter et al., 2008).

Limitations of the study include a relatively small sample size, which was dictated by the use of multiple stimulus types and ISIs, recording both MEG and fMRI data, and the resulting long recording sessions. Future studies with more subjects are needed to test if some of the differences that were non-significant in the current data could emerge as significant.

Functional roles of early cross-sensory activations are still incompletely understood. Audiovisual interactions for letters are clearly stronger after 300 ms than at these early latencies, and differentiate between matching and non-matching letter pairs

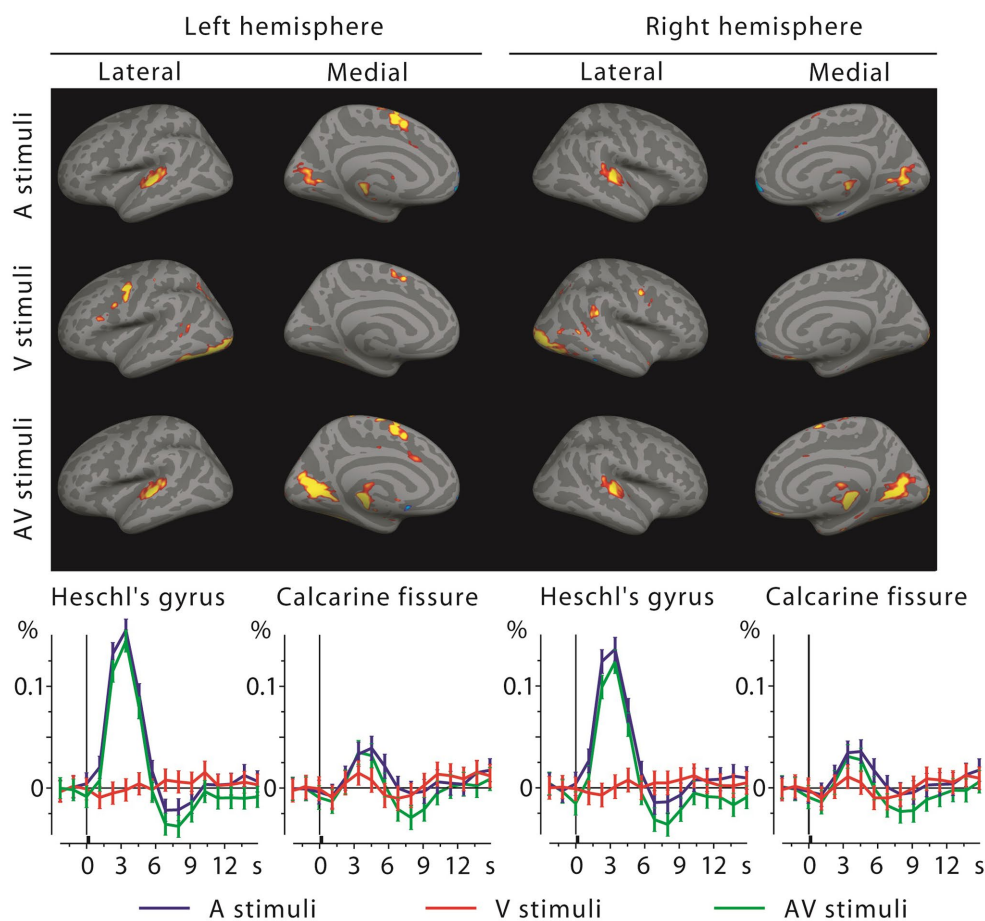


FIGURE 6

fMRI activations to auditory, visual, and audiovisual stimuli at the 4th time frame after stimulus onset (top, showing mainly positive activations at this frame) and the corresponding BOLD signal time courses with across-subjects SEM error bars from Heschl's gyri and calcarine fissures (bottom). Sensory-specific activations are strong for auditory letters but quite weak for visual letters (in V1, below visualization threshold in brain activation maps); cross-sensory responses are clear in Calcarine fissures bilaterally, but absent in the left and weak in the right Heschl's gyrus (see Discussion).

even later, after 400 ms; moreover, such activations are maximal in higher-order association areas such as STS (Raij et al., 2000). Plausibly, the early cross-sensory influences in low-order sensory areas may have a role for situations where synchrony requirements may be tight, and/or could facilitate later processing stages and reaction times by speeding up the exchange of signals between brain areas and enhancing top-down processing (Ullman, 1996; Bar et al., 2006; Raij et al., 2008; Sperdin et al., 2009). They may also be behaviorally particularly relevant in, e.g., situations where stimuli in one or more modalities are noisy (Jääskeläinen et al., 2011; Schepers et al., 2015; Bizley et al., 2016; Ahveninen et al., 2024). Overall, the fact that the onset latency differences between simpler stimuli and letters were small agrees with the view that the earliest cross-modal activations in sensory cortices reflect relatively automatic bottom-up processes (van Atteveldt et al., 2014; De Meo et al., 2015).

The observed cross-sensory onset latencies are, to our knowledge, the fastest reported for letter stimuli, and both these and sensory-specific latencies are consistent with those previously reported for simpler stimuli (Raij et al., 2010; Ahveninen et al., 2024). These findings contribute to understanding the timing and

potential anatomical pathways of early cross-sensory activations and interactions in sensory cortices in language processing. Further, the present quantification of millisecond-level cross-sensory conduction delays may enable future studies that manipulate effective connectivity between the stimulated brain areas (Hernandez-Pavon et al., 2022).

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Massachusetts General Hospital IRB. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

TR: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. F-HL: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Writing – review & editing. BL: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – review & editing. KL: Methodology, Software, Visualization, Writing – review & editing, Formal analysis, Investigation, Validation. TN: Methodology, Software, Writing – review & editing, Data curation, Formal analysis, Investigation, Project administration, Visualization. TW: Conceptualization, Methodology, Resources, Software, Writing – review & editing, Data curation, Visualization. MH: Methodology, Resources, Software, Writing – review & editing. JA: Conceptualization, Investigation, Methodology, Resources, Software, Validation, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by grants from the National Institutes of Health (R01NS126337, R01MH130490, R01NS048279, R01HD040712, R01NS037462, R01MH083744, R21EB007298, R21DC010060, P41RR14075, R01DC016915, R01DC016765, R01DC017991), National Center for Research Resources, Harvard Catalyst Pilot Grant/The Harvard Clinical and Translational Science Center (NIH UL1 RR 025758–02 and financial contributions from participating organizations), Sigrid Juselius Foundation, Academy of Finland, Finnish Cultural Foundation, National Science Council, Taiwan (NSC 98–2320-B-002-004-MY3, NSC 97–2320-B-002-058-MY3), and National Health Research Institute, Taiwan (NHRI-EX97-9715EC).

References

- Ahveninen, J., Lee, H., Yu, H., Lee, C., Chou, C., Ahlfors, S., et al. (2024). Visual stimuli modulate local field potentials but drive no high-frequency activity in human auditory cortex. *J. Neurosci.* 44:e0890232023. doi: 10.1523/JNEUROSCI.0890-23.2023
- Andres, A., Cardy, J., and Joannis, M. (2011). Congruency of auditory sounds and visual letters modulates mismatch negativity and P300 event-related potentials. *Int. J. Psychophysiol.* 79, 137–146. doi: 10.1016/j.ijpsycho.2010.09.012
- Bar, M., Kassam, K., Ghuman, A., Boshyan, J., Schmid, A., Dale, A., et al. (2006). Top-down facilitation of visual recognition. *Proc. Natl. Acad. Sci. USA* 103, 449–454. doi: 10.1073/pnas.0507062103
- Beer, A., Plank, T., and Greenlee, M. (2011). Diffusion tensor imaging shows white matter tracts between human auditory and visual cortex. *Exp. Brain Res.* 213, 299–308. doi: 10.1007/s00221-011-2715-y
- Beer, A., Plank, T., Meyer, G., and Greenlee, M. (2013). Combined diffusion-weighted and functional magnetic resonance imaging reveals a temporal-occipital network involved in auditory-visual object processing. *Front. Integr. Neurosci.* 7:5. doi: 10.3389/fnint.2013.00005
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Bizley, J., Maddox, R., and Lee, A. (2016). Defining auditory-visual objects: behavioral tests and physiological mechanisms. *Trends Neurosci.* 39, 74–85. doi: 10.1016/j.tins.2015.12.007
- Blau, V., Reithler, J., Van Atteveldt, N., Seitz, J., Gerretsen, P., Goebel, R., et al. (2010). Deviant processing of letters and speech sounds as proximate cause of reading failure: a

Acknowledgments

We thank John W. Belliveau, Valerie Carr, Sasha Devore, Deirdre Foxe, Mark Halko, Hsiao-Wen Huang, Yu-Hua Huang, Emily Israeli, Iiro Jääskeläinen, Natsuko Mori, Barbara Shinn-Cunningham, Mark Vangel, and Dan Wakeman for help.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of *Frontiers*, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnint.2024.1427149/full#supplementary-material>

functional magnetic resonance imaging study of dyslexic children. *Brain* 133, 868–879. doi: 10.1093/brain/awp308

Blomert, L. (2011). The neural signature of orthographic-phonological binding in successful and failing reading development. *NeuroImage* 57, 695–703. doi: 10.1016/j.neuroimage.2010.11.003

Brang, D., Towle, V., Suzuki, S., Hillyard, S., Di Tusa, S., Dai, Z., et al. (2015). Peripheral sounds rapidly activate visual cortex: evidence from electrocorticography. *J. Neurophysiol.* 114, 3023–3028. doi: 10.1152/jn.00728.2015

Burock, M., and 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. doi: 10.1002/1097-0193(200012)11:4<#x0003c;249::AID-HBM208#x0003e;3.0.CO;2-5

Cappe, C., Thut, G., Romei, V., and Murray, M. (2010). Auditory-visual multisensory interactions in humans: timing, topography, directionality, and sources. *J. Neurosci.* 30, 12572–12580. doi: 10.1523/JNEUROSCI.1099-10.2010

Chang, C., and Lin, C. (2011). LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2, 1–27. doi: 10.1145/1961189.1961199

Cohen, D., Schlapfer, U., Ahlfors, S., Hämäläinen, M., and Halgren, E. (2002). "New six-layer magnetically-shielded room for MEG" in 13th international conference on biomagnetism. eds. H. Nowak, J. Hauelsen, F. Giessler and R. Huonker ((VDE Verlag GmbH)), 919–921.

Cox, R., and Jesmanowicz, A. (1999). Real-time 3D image registration for functional MRI. *Magn. Reson. Med.* 42, 1014–1018. doi: 10.1002/(sici)1522-2594(199912)42:6<1014::aid-mrm4>3.0.co;2-f

- Dale, A. (1999). Optimal experimental design for event-related fMRI. *Hum. Brain Mapp.* 8, 109–114. doi: 10.1002/(SICI)1097-0193(1999)8:2/3<109::AID-HBM7>3.0.CO;2-W
- Dale, A., Liu, A., Fischl, B., Buckner, R., Belliveau, J., Lewine, J., et al. (2000). Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron* 26, 55–67. doi: 10.1016/s0896-6273(00)81138-1
- Dale, A., and 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. doi: 10.1162/jocn.1993.5.2.162
- De Meo, R., Murray, M., Clarke, S., and Matusz, P. (2015). Top-down control and early multisensory processes: chicken vs. egg. *Front. Integr. Neurosci.* 9:17. doi: 10.3389/fnint.2015.00017
- Desikan, R., Segonne, F., Fischl, B., Quinn, B., Dickerson, B., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* 31, 968–980. doi: 10.1016/j.neuroimage.2006.01.021
- Eckert, M., Kamdar, N., Chang, C., Beckmann, C., Greicius, M., and Menon, V. (2008). A cross-modal system linking primary auditory and visual cortices: evidence from intrinsic fMRI connectivity analysis. *Hum. Brain Mapp.* 29, 848–857. doi: 10.1002/hbm.20560
- Fischl, B., Salat, D., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., et al. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33, 341–355. doi: 10.1016/s0896-6273(02)00569-x
- Fischl, B., Sereno, M., Tootell, R., and Dale, A. (1999). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum. Brain Mapp.* 8, 272–284. doi: 10.1002/(sici)1097-0193(1999)8:4<272::aid-hbm10>3.0.co;2-4
- Fischl, B., Van Der Kouwe, A., Destrieux, C., Halgren, E., Segonne, F., Salat, D., et al. (2004). Automatically parcellating the human cerebral cortex. *Cereb. Cortex* 14, 11–22. doi: 10.1093/cercor/bhg087
- Foxe, J., Morocz, I., Murray, M., Higgins, B., Javitt, D., and 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. doi: 10.1016/s0926-6410(00)00024-0
- Foxe, J., and Schroeder, C. (2005). The case for feedforward multisensory convergence during early cortical processing. *Neuroreport* 16, 419–423. doi: 10.1097/00001756-200504040-00001
- Foxe, J., and 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. doi: 10.1007/s00221-001-0906-7
- Froyen, D., Willems, G., and Blomert, L. (2011). Evidence for a specific cross-modal association deficit in dyslexia: an electrophysiological study of letter-speech sound processing. *Dev. Sci.* 14, 635–648. doi: 10.1111/j.1467-7687.2010.01007.x
- Gaab, N., Gabrieli, J., and Glover, G. (2007a). Assessing the influence of scanner background noise on auditory processing. I. An fMRI study comparing three experimental designs with varying degrees of scanner noise. *Hum. Brain Mapp.* 28, 703–720. doi: 10.1002/hbm.20298
- Gaab, N., Gabrieli, J., and Glover, G. (2007b). Assessing the influence of scanner background noise on auditory processing. II. An fMRI study comparing auditory processing in the absence and presence of recorded scanner noise using a sparse design. *Hum. Brain Mapp.* 28, 721–732. doi: 10.1002/hbm.20299
- Ghazanfar, A., and Schroeder, C. (2006). Is neocortex essentially multisensory? *Trends Cogn. Sci.* 10, 278–285. doi: 10.1016/j.tics.2006.04.008
- Giard, M., and Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J. Cogn. Neurosci.* 11, 473–490. doi: 10.1162/089892999563544
- Goldenholz, D., Ahlfors, S., Hämäläinen, M., Sharon, D., Ishitobi, M., Vaina, L., et al. (2009). Mapping the signal-to-noise-ratios of cortical sources in magnetoencephalography and electroencephalography. *Hum. Brain Mapp.* 30, 1077–1086. doi: 10.1002/hbm.20571
- Gramfort, A., Luessi, M., Larson, E., Engemann, D., Strohmeier, D., Brodbeck, C., et al. (2014). MNE software for processing MEG and EEG data. *NeuroImage* 86, 446–460. doi: 10.1016/j.neuroimage.2013.10.027
- Hall, D., Haggard, M., Akeroyd, M., Palmer, A., Summerfield, A., Elliott, M., et al. (1999). “sparse” temporal sampling in auditory fMRI. *Hum. Brain Mapp.* 7, 213–223. doi: 10.1002/(sici)1097-0193(1999)7:3<213::aid-hbm5>3.0.co;2-n
- Hämäläinen, M., and Hari, R. (2002). “Magnetoencephalographic characterization of dynamic brain activation. Basic principles and methods of data collection and source analysis” in Brain mapping: The methods. ed. J. M. Aw Toga. 2nd edition ed (New York: Academic Press), 227–253. doi: 10.1016/B978-012693019-1/50012-5
- Hämäläinen, M., and Ilmoniemi, R. (1984). Interpreting measured magnetic fields of the brain: Estimates of current distributions. Helsinki, Finland: Helsinki University of Technology.
- Hämäläinen, M., and Ilmoniemi, R. (1994). Interpreting magnetic fields of the brain: minimum norm estimates. *Med. Biol. Eng. Comput.* 32, 35–42. doi: 10.1007/BF02512476
- Hennel, F., Girard, F., and Loenneker, T. (1999). “silent” MRI with soft gradient pulses. *Magn. Reson. Med.* 42, 6–10. doi: 10.1002/(sici)1522-2594(199907)42:1<6::aid-mrm2>3.0.co;2-d
- Henschke, J., Noesselt, T., Scheich, S., and Budge, E. (2015). Possible anatomical pathways for short-latency multisensory integration processes in primary sensory cortices. *Brain Struct. Funct.* 220, 955–977. doi: 10.1007/s00429-013-0694-4
- Herdman, A., Fujioka, T., Chau, W., Ross, B., Pantev, C., and Picton, T. (2006). Cortical oscillations related to processing congruent and incongruent grapheme-phoneme pairs. *Neurosci. Lett.* 399, 61–66. doi: 10.1016/j.neulet.2006.01.069
- Hernandez-Pavon, J., Schneider-Garces, N., Begnoche, J., Miller, L., and Raij, T. (2022). Targeted modulation of human brain interregional effective connectivity with spike-timing dependent plasticity. *NeuroModulation Technol. Neural Interface* 26, 745–754. doi: 10.1016/j.neurom.2022.10.045
- Jääskeläinen, I., Ahveninen, J., Andermann, M., Belliveau, J., Raij, T., and Sams, M. (2011). Short-term plasticity as a mechanism supporting memory and attentional functions. *Brain Res.* 1422, 66–81. doi: 10.1016/j.brainres.2011.09.031
- Kayser, C., and Logothetis, N. (2007). Do early sensory cortices integrate cross-modal information? *Brain Struct. Funct.* 212, 121–132. doi: 10.1007/s00429-007-0154-0
- Lankinen, K., Ahveninen, J., Jas, M., Raij, T., and Ahlfors, S. (2024). Neuronal modeling of cross-sensory visual evoked magnetoencephalography responses in the auditory cortex. *J. Neurosci.* 44:e119232024. doi: 10.1523/JNEUROSCI.1119-23.2024
- Liang, M., Mouraux, A., and Iannetti, G. (2013). Bypassing primary sensory cortices - a direct thalamocortical pathway for transmitting salient sensory information. *Cereb. Cortex* 23, 1–11. doi: 10.1093/cercor/bhr363
- Liu, A., Belliveau, J., and 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. doi: 10.1073/pnas.95.15.8945
- Lohse, M., Dahmen, J., Bajo, V., and King, A. (2021). Subcortical circuits mediate communication between primary sensory cortical areas in mice. *Nat. Commun.* 12:3916. doi: 10.1038/s41467-021-24200-x
- Macaluso, E. (2006). Multisensory processing in sensory-specific cortical areas. *Neuroscientist* 12, 327–338. doi: 10.1177/1073858406287908
- Macaluso, E., and Driver, J. (2005). Multisensory spatial interactions: a window onto functional integration in the human brain. *Trends Neurosci.* 28, 264–271. doi: 10.1016/j.tins.2005.03.008
- Martuzzi, R., Murray, M., Maeder, P., Fornari, E., Thiran, J., Clarke, S., et al. (2006). Visuo-motor pathways in humans revealed by event-related fMRI. *Exp. Brain Res.* 170, 472–487. doi: 10.1007/s00221-005-0232-6
- Martuzzi, R., Murray, M., Michel, C., Thiran, J.-P., Maeder, P., Clarke, S., et al. (2007). Multisensory interactions within human primary cortices revealed by BOLD dynamics. *Cereb. Cortex* 17, 1672–1679. doi: 10.1093/cercor/bhl077
- Molholm, S., and Foxe, J. (2005). Look ‘hear’, primary auditory cortex is active during lip-reading. *Neuroreport* 16, 123–124. doi: 10.1097/00001756-200502080-00009
- Molholm, S., Ritter, W., Javitt, D., and Foxe, J. (2004). Multisensory visual-auditory object recognition in humans: a high-density electrical mapping study. *Cereb. Cortex* 14, 452–465. doi: 10.1093/cercor/bhh007
- Molholm, S., Ritter, W., Murray, M., Javitt, D., Schroeder, C., and 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. doi: 10.1016/s0926-6410(02)00066-6
- Morrell, L. (1968). Sensory interaction: evoked potential observations in man. *Exp. Brain Res.* 6, 146–155. doi: 10.1007/BF00239168
- Murray, M., Molholm, S., Michel, C., Heslenfeld, D., Ritter, W., Javitt, D., et al. (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. doi: 10.1093/cercor/bhh197
- Musacchia, G., and Schroeder, C. (2009). Neuronal mechanisms, response dynamics and perceptual functions of multisensory interactions in auditory cortex. *Hear. Res.* 258, 72–79. doi: 10.1016/j.heares.2009.06.018
- Oosterhof, N., Connolly, A., and Haxby, J. (2016). CoSMoMVA: multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU octave. *Front. Neuroinform.* 10:27. doi: 10.3389/fninf.2016.00027
- Pekkola, J., Ojanen, V., Autti, T., Jääskeläinen, I., Mottonen, R., Tarkiainen, A., et al. (2005). Primary auditory cortex activation by visual speech: an fMRI study at 3 T. *Neuroreport* 16, 125–128. doi: 10.1097/00001756-200502080-00010
- Raij, T. (1999). Patterns of brain activity during visual imagery of letters. *J. Cogn. Neurosci.* 11, 282–299. doi: 10.1162/089892999563391
- Raij, T., Ahveninen, J., Lin, F., Witzel, T., Jääskeläinen, I., Letham, B., et al. (2010). Onset timing of cross-sensory activations and multisensory interactions in auditory and visual sensory cortices. *Eur. J. Neurosci.* 31, 1772–1782. doi: 10.1111/j.1460-9568.2010.07213.x
- Raij, T., Karhu, J., Kičić, D., Lioumis, P., Julkunen, P., Lin, F., et al. (2008). Parallel input makes the brain run faster. *NeuroImage* 40, 1792–1797. doi: 10.1016/j.neuroimage.2008.01.055

- Raij, T., Uutela, K., and Hari, R. (2000). Audiovisual integration of letters in the human brain. *Neuron* 28, 617–625. doi: 10.1016/s0896-6273(00)00138-0
- Schepers, I., Yoshor, D., and Beauchamp, M. (2015). Electro-corticography reveals enhanced visual cortex responses to visual speech. *Cereb. Cortex* 25, 4103–4110. doi: 10.1093/cercor/bhu127
- Schmitter, S., Diesch, E., Amann, M., Kroll, A., Moayer, M., and Schad, L. (2008). Silent echo-planar imaging for auditory fMRI. *MAGMA* 21, 317–325. doi: 10.1007/s10334-008-0132-4
- Schroeder, C., and 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. doi: 10.1016/s0926-6410(02)00073-3
- Schroeder, C., and Foxe, J. (2005). Multisensory contributions to low-level, 'unisensory' processing. *Curr. Opin. Neurobiol.* 15, 454–458. doi: 10.1016/j.conb.2005.06.008
- Schroeder, C., Lindsley, R., Specht, C., Marcovici, A., Smiley, J., and Javitt, D. (2001). Somatosensory input to auditory association cortex in the macaque monkey. *J. Neurophysiol.* 85, 1322–1327. doi: 10.1152/jn.2001.85.3.1322
- Schroeder, C., Smiley, J., Fu, K., McGinnis, T., O'Connell, M., and Hackett, T. (2003). Anatomical mechanisms and functional implications of multisensory convergence in early cortical processing. *Int. J. Psychophysiol.* 50, 5–17. doi: 10.1016/s0167-8760(03)00120-x
- Schwarzbauer, C., Davis, M., Rodd, J., and Johnsrude, I. (2006). Interleaved silent steady state (ISSS) imaging: a new sparse imaging method applied to auditory fMRI. *NeuroImage* 29, 774–782. doi: 10.1016/j.neuroimage.2005.08.025
- Sperdin, H., Cappe, C., Foxe, J., and Murray, M. (2009). Early, low-level auditory-somatosensory multisensory interactions impact reaction time speed. *Front. Integr. Neurosci.* 3:2. doi: 10.3389/neuro.07.002.2009
- Talsma, D., Doty, T., and Woldorff, M. (2007). Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? *Cereb. Cortex* 17, 679–690. doi: 10.1093/cercor/bhk016
- Teder-Sälejärvi, W., McDonald, J., Di Russo, F., and 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. doi: 10.1016/s0926-6410(02)00065-4
- Ullman, S. (1996). "Sequence seeking and counter streams: A model for visual cortex" in High-level vision: Object recognition and visual cognition. ed. S. Ullman (Cambridge MA: MIT Press).
- Van Atteveldt, N., Formisano, E., Goebel, R., and Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron* 43, 271–282. doi: 10.1016/j.neuron.2004.06.025
- Van Atteveldt, N., Murray, M., Thut, G., and Schroeder, C. (2014). Multisensory integration: flexible use of general operations. *Neuron* 81, 1240–1253. doi: 10.1016/j.neuron.2014.02.044
- Yang, Y., Engelen, A., Engelen, W., Xu, S., Stern, E., and Silbersweig, D. (2000). A silent event-related functional MRI technique for brain activation studies without interference of scanner acoustic noise. *Magn. Reson. Med.* 43, 185–190. doi: 10.1002/(sici)1522-2594(200002)43:2<185::aid-mrm4>3.0.co;2-3