Visual stimuli modulate local field potentials but drive no high-frequency activity in human auditory cortex

Jyrki Ahveninen [1,2*], Hsin-Ju Lee [4, 5*], Hsiang-Yu Yu [6,9], Cheng-Chia Lee [7,9], Chien-Chen Chou [6,9], Seppo P. Ahlfors [1,2], Wen-Jui Kuo [8], Iiro P. Jääskeläinen [3,10 §], Fa-Hsuan Lin [3, 4, 5 §]

[1] Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Charlestown, MA, USA
[2] Department of Radiology, Harvard Medical School, Boston, MA, USA
[3] Brain and Mind Laboratory, Department of Neuroscience and Biomedical Engineering, Aalto University School of Science, Espoo, Finland
[4] Physical Sciences Platform, Sunnybrook Research Institute, Toronto, ON, Canada
[5] Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada
[6] Department of Epilepsy, Neurological Institute, Taipei Veterans General Hospital, Taipei, Taiwan
[7] Department of Neurosurgery, Neurological Institute, Taipei Veterans General Hospital, Taipei, Taiwan
[8] Institute of Neuroscience, National Yang Ming Chiao Tung University, Taipei, Taiwan
[9] School of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan
[10] International Laboratory of Social Neurobiology, Institute of Cognitive Neuroscience, Higher School of Economics, Moscow, Russia

* Authors JA and HJL contributed equally; [§] IPJ and FHL are joint senior authors.

Corresponding author:

Jyrki Ahveninen, Ph.D.

CNY 149, 13th St.

A.A. Martinos Center for Biomedical Imaging,

Department of Radiology,

Massachusetts General Hospital,

Charlestown, MA 02129

jahveninen@mgh.harvard.edu

Tel: 617 726 6584, Fax: (617) 726-7422

Abstract

Neuroimaging studies suggest cross-sensory visual influences in human auditory cortices (AC). Whether these influences reflect active visual processing in human ACs, which drives neuronal firing and concurrent broadband high-frequency activity (BHFA; >70 Hz), or whether they merely modulate sound processing is still debatable. Here, we presented auditory, visual, and audiovisual stimuli to 16 participants with stereo-EEG depth electrodes implanted near ACs for presurgical monitoring. Anatomically normalized group analyses were facilitated by inverse modeling of intracranial source currents. Pattern cross-correlation analyses suggested cross-sensory visual influences, lagging auditory responses by ~50 ms in ACs. Visual stimuli also modulated the phase of intrinsic low-frequency oscillations and triggered 15–30-Hz event-related desynchronization in ACs. However, BHFA, a putative correlate of neuronal firing, was not significantly increased in ACs after visual stimuli, not even when they coincided with auditory stimuli. Intracranial recordings demonstrate cross-sensory modulations, but no indication of active visual processing in human ACs.

Introduction

Many previous studies support the view that cross-sensory visual information influences the function of human AC, possibly involving even primary areas [1-4]. The exact functional significance of these early cross-sensory influences, however, remains still unclear. A more conservative theory suggests that cross-sensory influences play a modulatory role in early sensory areas including ACs, to enhance relevant and suppress irrelevant ascending inputs, akin to the effects of selective attention. This hypothesis is supported by neurophysiological evidence that the effects of unimodal visual stimuli on ACs are limited to subthreshold synaptic influences [5-7]. For example, laminar extra-cellular recordings in the non-human primate (NPH) AC suggest that when presented alone, cross-sensory visual stimuli modulate the phase of neuronal oscillations but do not evoke increased neuronal firing detectable in the analysis of multiunit activity [8]. Similarly, systematic neurophysiological recordings in rodents suggest that the prevalence of individual neurons activated to other than sensory-specific inputs is very low in sensory areas including AC [9].

Classically, integration across senses has been thought to be mediated by neurons that show increased firing activity to concurrent multisensory inputs. These integrative effects were originally quantified as *multisensory interactions* (MSI), changes in firing rates of individual neurons, which are non-additive and which are strongest for faint or noisy stimuli [10]. For example, MSIs between a faint sound and spatially and temporally coinciding visual [10] or tactile [11] stimulus result in activations stronger than the sum of the respective unimodal activations. Neurophysiological evidence for such MSI effects has been obtained from higher areas [12, 13], but whether the human ACs contains integrative neurons that increase their firing rate and synchrony when sounds are coupled with visual stimuli is still uncertain. One previous NPH model suggested that despite the significant effects observable in local field potential (LFP) responses, measures of firing activity were concurrently suppressed rather than increased [6]. Effects such as these would support the more conservative hypothesis that cross-sensory influences modulate but do not activate neurons in ACs.

However, many previous findings have been interpreted to provide support for a more provocative hypothesis that multisensory processing occurs even in early human AC [14-16]. This hypothesis has gained most direct support from neurophysiological studies in animal models, including single-unit recordings in non-human primates [17] and other mammals [9, 18, 19], which provide evidence for cross-sensory firing activity in ACs.

According to one of these studies, which was conducted in the ferret model, cross-sensory stimuli drive single-unit firing patterns that convey information of the properties visual stimuli in ACs [18]. Similar findings emerged in a recent study that suggested firing activity to visual stimuli in the deep infragranular layers of mouse AC [20]. These findings would suggest that instead of being only modulated by cross-sensory influences, AC neurons may actively process visual information.

Whether influences such as those seen in animal models occur also in the human AC is difficult to examine using conventional non-invasive neuroimaging paradigms. The bulk of human neuroimaging evidence of AC responses triggered by unimodal cross-sensory stimuli consists of studies using MEG/EEG [3, 21] and fMRI [4, 14], which offer limited means for making inferences of neuronal mechanisms. At the same time, the few previous studies using intracranial recordings to investigate cross-sensory influences of simple non-verbal stimuli in ACs have typically been limited to relatively small samples of participants (N=3-8) and focused on individual-level statistical approaches [16, 22], which may limit the generalizability of results.

Here, we addressed the question whether human ACs are not only modulated, but also activated by cross-sensory visual stimuli. To this end, we employed intracranial stereo-EEG (SEEG) recordings in a relatively large sample of patients (N=16) who had depth electrodes implanted in or near auditory cortical areas for preoperative monitoring. This methodology provides direct measurements of LFPs from the neuronal tissue, to quantify broadband high-frequency activity (BHFA, above 70 Hz) activity. in contrast to gamma band oscillations that are also visible to MEG and EEG, BHFA signals are believed to be a reliable correlates of local spiking activity [23-26]. Another benefit of using SEEG data is that the recording contacts of depth electrodes extend across the depth of the superior temporal plane help reveal the patterns of auditory vs. cross-sensory responses at a much greater detail than what is possible with non-invasive recordings, or even with subdural electrocorticography (ECoG). A challenge in previous human SEEG studies of cross-sensory modulation of human ACs has been that the clinically determined anatomical implantation plans vary between participants, thereby complicating robust hypothesis-testing at the group level [16, 27]. Here, to facilitate group analyses, we complemented traditional "electrode-space" analyses with our recently introduced surface-based source modeling technique that estimates the neuronal activity in the anatomically normalized cortical "source space" [28].

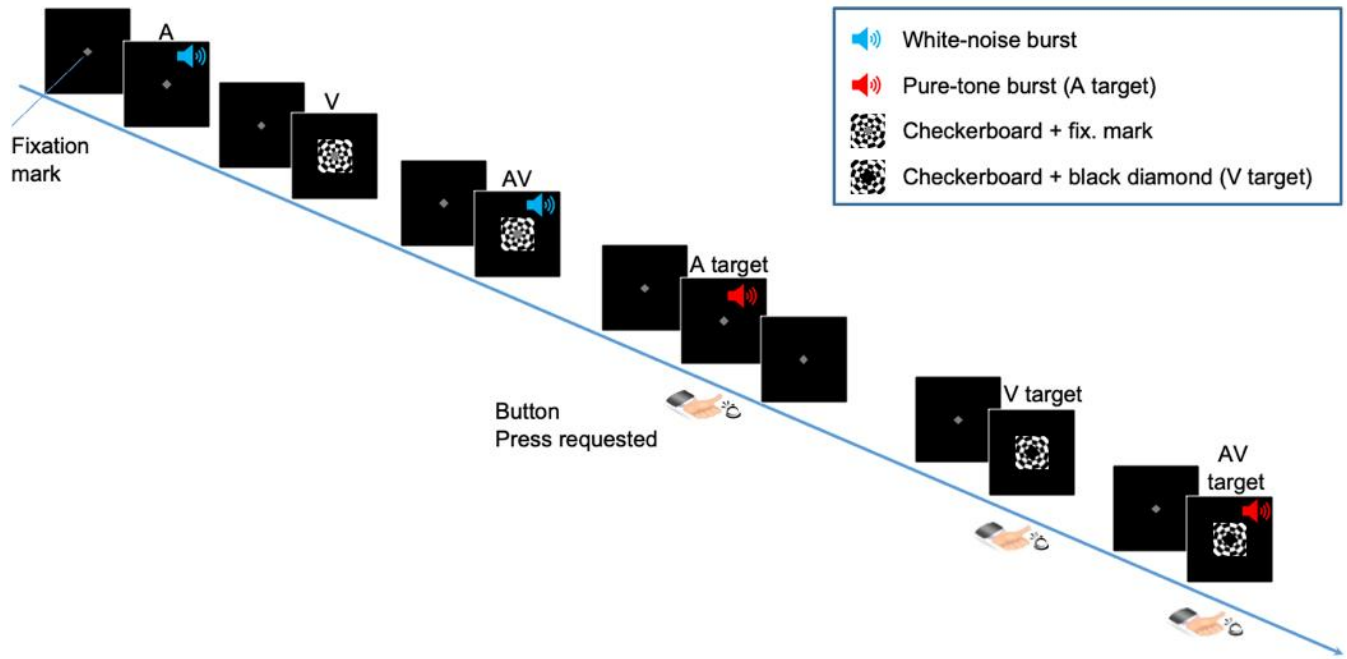


Figure 1. Stimuli and task. Participants were presented with simple A, V, and AV stimuli. In an oddball design, they were asked to press a button upon hearing a rare A, V, or AV target stimulus among frequent non-targets. The non-targets were A noise bursts, V checkerboard patterns, or their AV combinations. The A targets were A pure tones, V targets diamond shapes embedded within a checkerboard, or their AV combinations.

Results

To investigate cross-sensory influences in AC, we presented auditory (A), visual (V), and audiovisual (AV) stimuli to 16 participants with stereo-EEG (SEEG) depth electrodes implanted near ACs for presurgical monitoring. The stimuli were presented in a randomized order, at a 1.2–2.8 s jittered inter-stimulus interval (ISI; mean 2 s) during SEEG recordings (**Fig. 1**). Using this design, we found strong evidence of subthreshold modulatory effects, reflected as "phase resetting" of ongoing intrinsic low-frequency oscillations and enhanced event-related desynchronizations (ERDS) of 15-30 Hz beta oscillations in ACs. Furthermore, our analyses suggested significant pattern similarity to Laplacian referenced LFP responses to A and V stimuli across the contacts of SEEG depth electrodes passing through superior temporal areas (i.e., "sT" electrodes). According to our cross-correlation analyses, these visual LFP influences lagged intra-modal auditory LFPs by about 50 ms. However, although the participants significantly benefited from combining information from the two modalities in terms of the behavioral reaction times (RT), there was little evidence of enhanced BHFA to V stimuli, whether presented alone or in combination with A stimuli in our group analyses.

Behavioral results

All 16 participants were able to perform the task according to the instruction: On average, they were able to detect almost all targets in auditory (A), visual (V), and audiovisual (AV) conditions, with no statistically significant differences in proportions of correct responses (PCR) between the conditions. The mean ± standard error (SEM) values of PCR for A, V, and AV target stimuli were PCR_A = 0.92 ± 0.04; PCR_V = 0.93 ± 0.04 and PCR_{AV} = 0.94 ± 0.03, respectively. The high PCR value for the unimodal V targets provides evidence that the participants were able to focus their gaze according to the instruction. Evidence for significant advantages of combining A and V stimuli were found in RTs to correctly detected target stimuli. The mean ± SEM values of RT for A, V, and AV conditions were RT_A = 668 ± 24 ms, RT_V = 575 ± 19 ms, and RT_{AV} = 546 ± 24 ms, respectively. The RTs were significantly faster to AV targets than to A targets (t_{15} = -5.6, $p < 0.001$; Cohen's d= -1.5), as well as to AV targets than to V targets (t_{15} = -3.0, $p < 0.01$; d=-0.8).

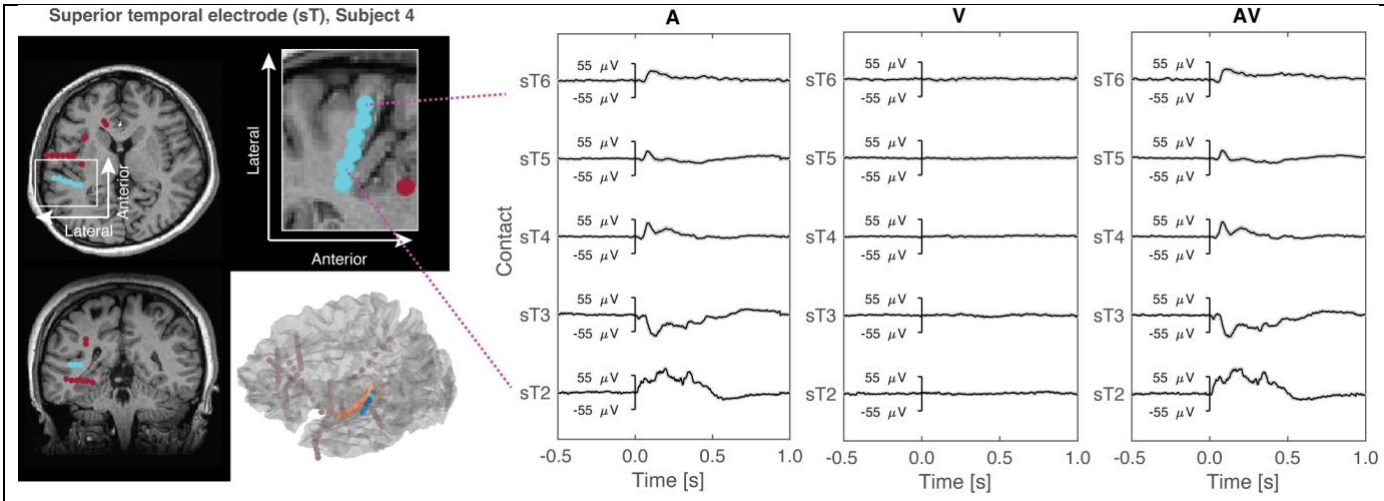


Figure 2. Intracranial local field potential (LFP) data data in a single participant. The locations of the superior temporal (sT) electrode contacts are shown on the left. Trial-averaged response waveforms are shown with the standard error of mean (SEM) across the trials, depicted by the shaded gray area. Note that in this participant, the SEM was very low, indicating the exceptional signal-to-noise ratio (SNR) of SEEG recordings. The data show a bipolar derivation with each contact being referenced to its nearest medial neighbor. The trial-averaged LFP traces of the rest of the participants included in the LFP pattern similarity analysis are shown in **Supplementary Figs. S1, S2**.

Pattern similarity of LFPs to auditory vs. visual stimuli

We compared auditory-cortical LFPs to A and V stimuli using a multivariate pattern analysis across the contacts in sT depth electrodes. This strategy was chosen because in comparison intramodal A inputs, responses to cross-sensory V stimuli are weak, as demonstrated by the trial-averaged LFP data shown in **Fig. 2, Suppl. Fig. S1** and **Suppl Fig**. **S2**. Moreover, pattern similarity analyses are suitable for finding commonalities in brain activations whose exact anatomical substrate is individually varying [29], as is the case here due to the clinically-driven individual differences of SEEG implantation plans. Given the Laplacian referencing strategy, only subjects with more than four adjacent contacts with acceptable signal quality in the sT electrode were included in this analysis (N=13 subjects). As shown in **Fig 3a**, which depicts the data of a single participant, the LFP pattern to V stimuli shares marked resemblance to those generated to A stimuli, despite the differences in the response magnitudes between the two conditions. Moreover, LFPs to V stimuli appear to originate from quite same medial Heschl's gyrus (HG) than the LFP responses to A stimuli, that is, in or near the primary AC. Analogous similarities emerged also in several other participants whose individual LFP patterns are shown in **Suppl. Fig. S3**.

To quantify the similarity of V vs. A activation patterns in AC at the group level, we compared the respective spatiotemporal LFP patterns by using Fisher-transformed Kendall's tau correlation coefficients calculated between the LFPs to A and V stimuli (**Fig. 3b**). These correlations were calculated across all combinations of 500-ms time windows of before and after the stimulus onsets, within and across the stimulus modalities. The average of transformed correlation coefficients involving the pre-stimulus baseline of A and/or V conditions was used as a "baseline", to mitigate spurious effects stemming from the similarity of noise levels in different contacts. It is worth noting that all correlations across A and V trials reflect correspondences between fully independent samples. The Kendal tau values computed between A and V post-stimulation periods (0-500 ms) was significantly larger than the "baseline" determined based all other combinations across pre- and post-stimulus time windows ($t(12)=3.3$, $p=0.003$, $d=0.91$). This result provides group-level evidence that V stimuli modulate the same aspect of superior temporal cortex than that activated by simple A stimuli alone, that is, the likely location of human early AC.

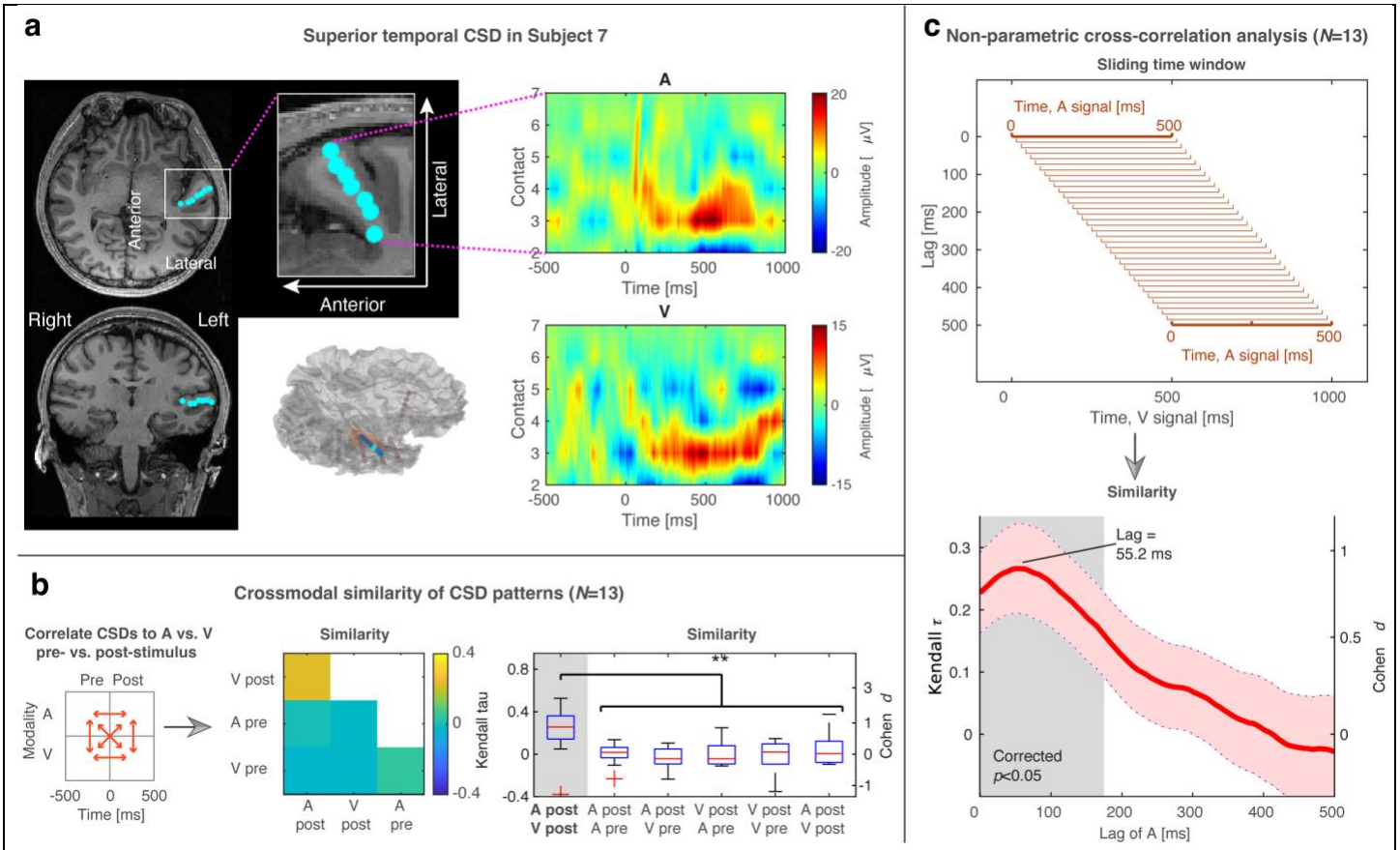


Figure 3. Pattern similarity analysis of cross-sensory modulation of human AC LFP signals. (**a**) Superior temporal responses to auditory (A) and visual (V) stimuli in a single participant. Although being weaker, the Laplacian-referenced LFPs to cross-sensory V stimuli showed marked similarity to that generated by A stimuli. The cyan dots demonstrate the locations of the contacts projected to one MRI slice. The outermost connection is not shown. (**b**) Group analysis of crossmodal similarity of LFP patterns across sT electrode contacts. The nonparametric Kendall correlation coefficient was computed across 500-ms periods from before ("pre") and after ("post") the stimulus onset. The correlation was significantly higher between A and V post-stimulus periods than between any of the other combinations. (**c**) Time-window-based non-parametric cross-correlation analysis of A vs. V pattern similarity. In a one-second sliding window analysis, the LFP pattern to A stimuli was compared to increasingly later parts of the LFPs to V stimuli. The cross-correlation peaked when the A signal was lagged by 55.2 ms. Taken together, these results suggest that cross-sensory V stimuli modulate the population of neurons that also generates LFP responses to simple A stimuli, at a lag of several tens of milliseconds. The unit scales in the panels **b** and **c** refer to Kendall $\tau = \tanh(\bar{z})$ where $\bar{z}$ is the group average of Fisher-transformed τ values in individual participants, converted to Cohen d as described in [30].

To estimate the relative timing of cross-sensory influences in auditory areas, we used a modified version of cross-correlation analysis of LFPs to A vs. V stimuli (**Fig. 3c**). In a sample-by-sample fashion, we correlated the

LFP pattern generated from 0 to 500 ms to A stimuli with increasingly later aspects of the LFP pattern generated to the V stimulus. The group average of the Kendall tau correlation coefficient peaked when the A time series was lagged by 55.2 ms relative to the V time series ($t(12)=3.8$, FDR-adjusted $p=0.012$; $d=0.89$ [30]).

Time-frequency representation of SEEG power

We utilized two strategies to allow hypothesis testing at the group level:

(**i**) For each subject, we determined the electrode contacts that were closest to the centroids of three cortical Human Connectome Project (HCP) labels across the auditory hierarchy (**Figs. 3a and 4a**; **Suppl. Fig. S4**). Our primary focus in the electrode-level analyses was the contact closest to the HCP label "early auditory cortex" (AC) (**Figs. 3b, 4b**). The group analysis results regarding contacts closest to the HCP labels "auditory association cortex" (AAC) and "temporo-parieto-occipital junction" (TPOJ) are included as Supplementary Material (**Suppl. Fig. S4**).

(**ii**) A limitation of the electrode-contact analysis is that the recording sites, which were determined by clinical reasons only, differ between the participants. Therefore, to allow for anatomically normalized analyses across the auditory processing hierarchy, we used our recently developed SEEG source modeling approach [28]. Hypothesis testing was conducted using LME models analogous to those used for electrode-contact analyses in the cortical ROIs AC, AAC, and TPOJ (**Figs. 3c and 4c)**. z

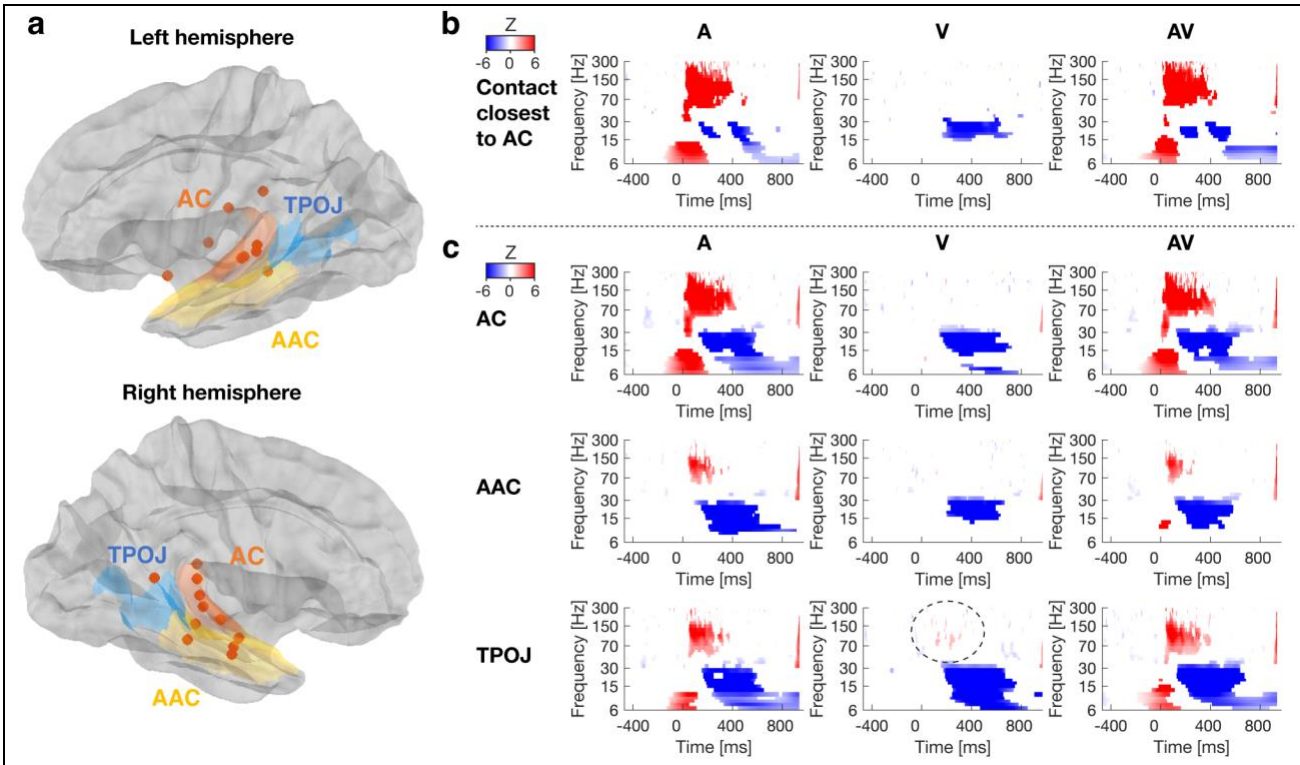


Figure 4. Group analysis of power TFRs. (**a**) Analysis sites in the standard brain representation. Each subject's SEEG contact that was closest to AC is marked with an orange dot. The orange, yellow, and blue labels show the three ROIs used for the source-modeling analysis, determined based on the HCP atlas. (**b**) TFRs to A, V, and AV stimuli in the electrode contact that was in each subject closest to AC. Statistically significant BHFA (>70 Hz, FDR-adjusted p<0.05) emerged to A and AV stimuli only. Cross-sensory V stimuli suppress beta (15-30 Hz) power 200 ms after the stimulus onset. (**c**) ROI analysis of SEEG source estimates. Consistent with the electrode-contact analysis, cross-sensory V triggered suppression of beta and theta/alpha (6-15 Hz) power across the studied ROIs. Traces of weak but significant (FDR-adjusted *p*<0.05) enhancement of BHFA was observed only in TPOJ (encircled). The TFRs depict group averages of the effect size *z* in individual subjects, masked to time/frequency elements with statistically significant effects (FDR-adjusted *p*<0.05) [31]. Individual subjects' frequency-wise *z* values were determined by subtracting the mean baseline power from each TFR element and then dividing them by the baseline SD. Data from both hemispheres were included a single linear mixed effects (LME) model to determine the statistical significance.

TFRs of baseline-normalized signal power values to A, V, and AV stimuli in these contacts of interest and the respective ROIs were then entered to linear mixed effects (LME) models comparing to the pre-stimulus baseline and determining the differences between AV-(A+V) AV-A, and AV-V contrasts (**Figs. 3, 4**). These LME models assessed the statistical significance of the intercept while controlling for the fixed effect of implanted hemisphere and the random effect of subject identity at all time and frequency instances, with the p-values corrected for multiple comparisons using the false discovery rate (FDR) procedure of [31]. The critical p value was determined jointly across all conditions/contrasts and contacts-of-interest at 0-500 ms and 6-250 Hz.

Stimulus condition effects: In the contact closest to the HCP label AC, TFRs to A and AV stimuli showed a characteristic event-related pattern where statistically significant (FDR-adjusted $p<0.05$) early power increase (or "event-related synchronization", ERS) at the 6-15 Hz theta/alpha range is followed by a beta-range (15-30 Hz) power suppression (or "event-related desynchronization", ERDS; **Fig. 4b**). This theta/alpha-beta ERS/ERDS pattern was coupled with a significant (FDR-adjusted $p<0.05$) sustained increase of BHFA. Interestingly, the beta-range ERDS was equally clear to V stimuli alone, suggesting statistically significant (FDR-adjusted $p<0.05$) cross-sensory modulation of beta-range oscillations in or near human ACs, which started around 200 ms after the stimulus onset. However, no evidence of increases of BHFA, a putative marker of underlying firing activity, were found in the contact closest to AC. The results for contacts closest to the HCP labels "auditory association cortex" (ACC) and "temporo-parieto-occipital junction" (TPOJ) are described in the Supplementary Material (**Suppl. Fig. S4**).

The A, V, and AV stimulus-related power TFRs of each individual participant's contact closest to AC are shown in **Supplementary Fig. S5**. In this individual-level analyses, the single-trial TFR power was compared to pre-stimulus baseline using paired *t*-tests across all trials. The purpose of comparisons against the baseline was also to normalize the TFRs across frequencies to mitigate the 1/f trend of signal power. To avoid false negatives in our interpretation of the lack of BHFA to V stimuli in ACs, we used no correction for multiple comparisons in these specific single-participant analyses. Despite the inclusive criterion, in the electrode-contact-space, evidence of more robust increase of visual BHFA, deemed qualitatively different from pre-stimulus noise, was found in only one participant, Subject 8. (**Suppl. Fig. S5**). However, in this participant, the contact that was in the 3D volume closest to early AC appears to be, in fact, located closer to the fundus of STS (a classic multisensory area) than AC, per se (**Suppl. Fig. S6**).

Results of our anatomically normalized group ROI source-modeling analysis agreed with the SEEG contact analysis (**Fig. 4c**). At the lower frequency range, responses to A and AV stimuli were characterized by an early theta/alpha ERS (<200 ms, FDR-adjusted $p<0.05$), followed by a longer-lasting beta ERDS (<30 Hz, FDR-adjusted $p<0.05$). Consistent with the electrode-contact analysis, this later lower-frequency ERDS effect was also prominently significant to V stimuli alone (FDR-adjusted $p<0.05$), constituting the strongest putatively cross-sensory influence across the auditory hierarchy in the present study. As for the BHFA range, statistically significant (FDR-corrected $p<0.05$) power increases were observed to A and AV stimuli alone in AC, AAC, and

TPOJ areas. No significant increases of cross-sensory BHFA were observed to V stimuli in AC or AAC. Traces of significantly increased BHFA to V stimuli were, however, observed in the area TPOJ (FDR-adjusted $p<0.05$). **Supplementary Figure S7** shows the corresponding source modeling results in each individual subject.

Contrast between stimulus conditions: The group analysis of TFRs from the contact closest to AC suggested significant (FDR-adjusted $p<0.05$) superadditive MSI effects, determined from the AV-(A+V) contrast, at the beta range (**Fig. 5b**). Other than the confirmatory contrast between AV and V conditions, where power changes to a stimulus containing auditory information vs. no auditory information are non-surprising, there was no evidence that the simultaneous presentation of A and V stimuli increases BHFA in the contact closest to AC. Similar to the electrode-contact analysis, the results of source modeling ROI analyses suggested significant (FDR-adjusted $p<0.05$) superadditive MSI effects in AC, as well as in AAC and TPOJ, which occurred at the theta and beta ranges. In the AV vs. A contrast, we also observed significant increases of early high alpha activity and subsequent beta ERDS effects. Besides the anticipated BHFA effects in the AV-V contrast, the only evidence of increased BHFA was in the AV-A contrast in TPOJ. In particular, no other BHFA effects were observed to AV-(A+V) or AV-A contrast. Results for the condition contrasts in the electrode contacts closest to AAC and TPOJ are shown in **Supplementary Figure S8**.

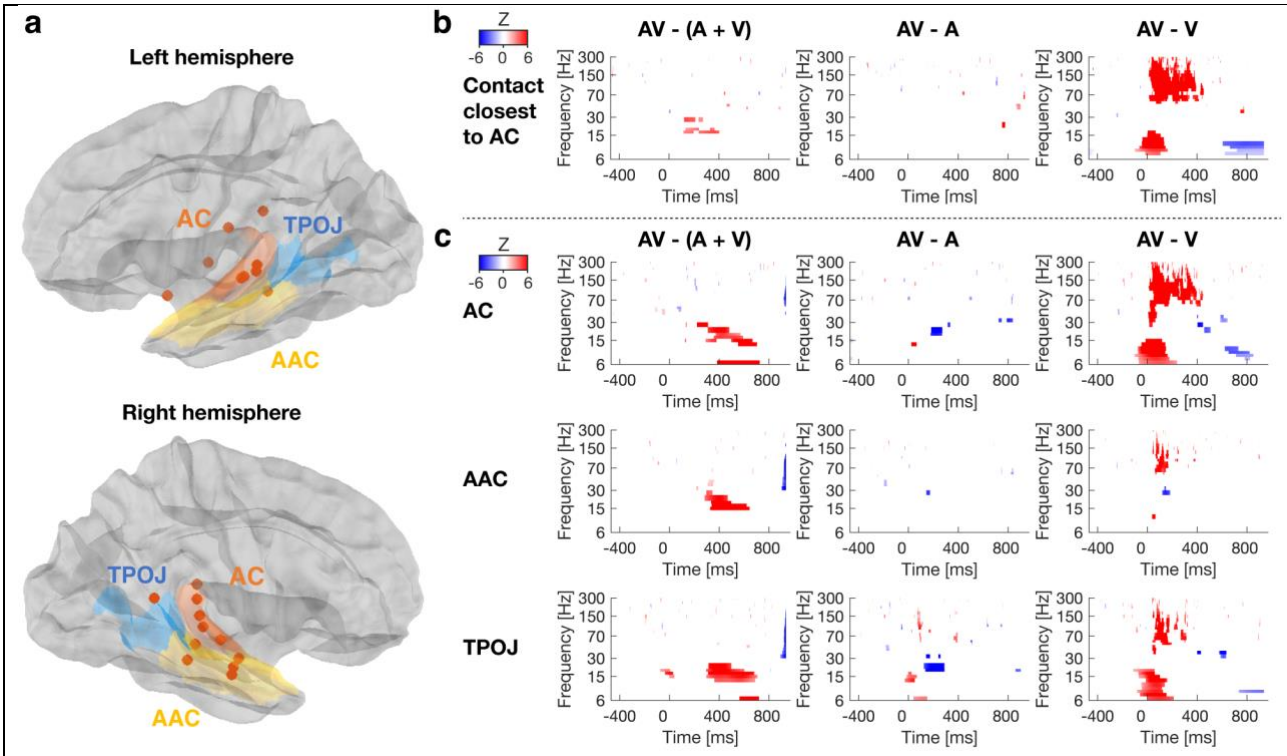


Fig. 5. Group analysis of power TFR contrasts between conditions. (**a**) Analysis sites in the standard brain representation. Each subject's SEEG contact that was closest to AC is marked with an orange dot. The orange, yellow, and blue labels show the three ROIs used for the source-modeling analysis, determined based on the HCP atlas. (**b**) TFR contrasts in the electrode contact closest to AC. Indices of statistically significant (FDR-adjusted p<0.05) multisensory interactions (MSI) occur at the beta range in the AV-(A+V) contrast. The contrast AV-V showed the anticipated enhancement of BHFA and early-latency theta/alpha in AC. (**c**) ROI analysis of source estimates. Indices significant of low-frequency MSIs were observed in all ROIs (FDR-adjusted p<0.05). In addition, beta power was significantly suppressed in AC and TPOJ in the AV-A contrast. Apart from the anticipated effects in the AV-V contrast, there was no evidence of effects at the BHFA range. The TFRs depict group averages of the baseline-normalized power, determined identically to thd data in **Fig. 4**, masked to time/frequency elements with statistically significant effects (FDR-adjusted *p*<0.05) [31]. Data from both hemispheres were included a single linear mixed effects (LME) model.

Intracranial inter-trial phase consistency

In the light of previous studies in non-human primates [8], we tested whether cross-sensory stimuli affect the phase of ongoing oscillations in human AC. The results suggest significant increases of low-frequency inter-trial phase consistency (ITPC) after V stimuli, both in the electrode contact closest to AC (**Fig. 6b**) and in all three source-modeling ROIs across the auditory processing hierarchy (FDR-adjusted p<0.05) (**Fig. 6c**). In the contact analysis, this ITPC effect was significant already about 100 ms after the onset of V stimuli. Most interestingly, the cross-sensory ITPC increases across the auditory processing hierarchy occurred at time/frequency windows

where there was no evidence for concurrent increases of TFR power. ITPC values in the electrode contacts closest to AAC and TPOJ are shown in **Suppl. Fig. S9**. The results of ITPC contrasts between conditions are shown in Supplementary Material (**Suppl. Figs. S10, S11**).

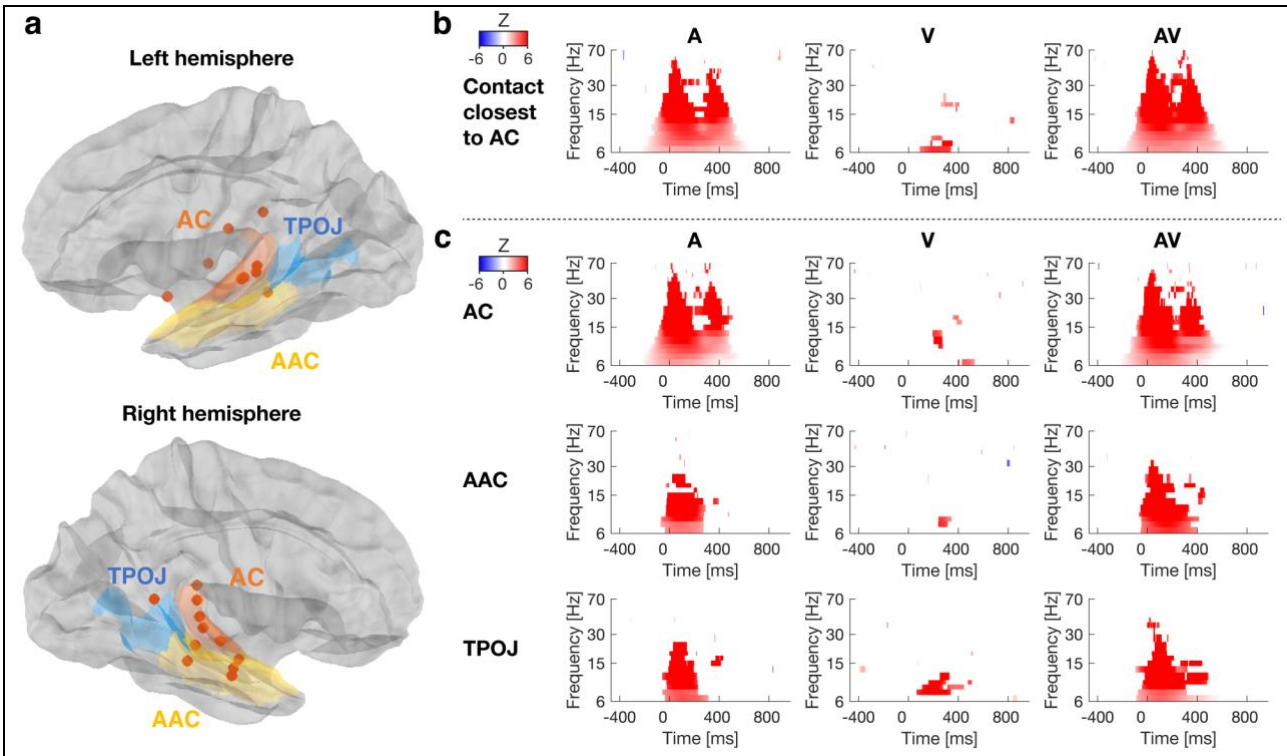


Fig. 6. Group analysis of ITPC. (**a**) Analysis sites in the standard brain representation. Each subject's SEEG contact closest to AC is marked with an orange dot. The orange, yellow, and blue labels show the HCP-based source-modeling ROIs. (**b**) Normalized ITPC in the electrode contact closest to AC. In addition to the strong ITPC effects to A and AV stimuli, there was also evidence of significant increase of low-frequency ITPC to V stimuli (FDR-adjusted $p<0.05$). These effects started less than 200 ms after the onset of the V stimulus. Most interestingly, the cross-sensory ITPC effect occurred in time/frequency windows where there was no evidence of power increases to V stimuli, in line with the hypothesis that cross-sensory influences could "entrain" intrinsic oscillations of AC [8]. (**c**) ROI-based group analysis of normalized ITPC values. In addition to responses to A and AV stimuli, there was evidence for significantly increased ITPC to V stimuli across the three ROIs. The data also show the expected early increases of ITPC to A and AV stimuli. Group averages of normalized ITPC values have been masked to time/frequency elements where the effects were statistically significant (FDR-adjusted $p<0.05$).

Discussion

Here, we used direct intracranial depth-electrode recordings in 16 human participants to compare two competing theories of how cross-sensory influences take place in and near the AC. To minimize the account of higher semantic or linguistic feedback influences, we used simple noise-burst sounds and "checkerboard" visual stimuli that lack strong auditory associations. In the frequency domain, visual stimuli modulated ITPC of low-frequency oscillations in AC without concurrent increases in signal power, suggesting cross-sensory phase resetting of intrinsic oscillatory activity [8]. Visual stimuli also resulted in a later beta-range ERDS effect in ACs. Similarity analyses of spatiotemporal LFP patterns provided further evidence of cross-sensory modulations of AC activity and suggested an about 50 ms lag between the visual and auditory LFP patterns in superior temporal cortex. However, no robust evidence of visually driven BHFA, a putative correlate of local neuronal firing [24-26], was found in ACs. No significant non-additive MSI effects (*i.e.*, AV ≠ A+V), which would reflect multisensory convergence in the classical sense [32], were found at the BHFA range either. MSI effects in ACs were concentrated at the beta range and below and these lower-frequency effects appeared to occur earlier in higher (TPOJ) than lower order (AC) areas along the auditory processing pathway.

The significant increase in ITPC, which resembles the results of an earlier subdural ECoG study in three human participants [22], could reflect transient modulations of the phase of intrinsic oscillations in AC. Evidence for this phenomenon was originally found in monkey studies of Peter Lakatos et al. [8] who investigated cross-sensory influences of somatosensory stimuli in primary AC. Lakatos et al. proposed that phase modulations, which occur without concurrent increases in the signal power, could be explained by phase resetting of ongoing intrinsic oscillations of AC. Such phase-resetting could, for example, help ensure that concurrent auditory inputs arrive to AC during at an optimal high-excitability phase, to amplify weak inputs during noisy listening conditions [8, 33, 34]. Support for this idea was provided by a recent intracranial EEG study, which suggested that AC neurons track the phase of visual speech and that this phase resetting amplifies the cortical representation of speech signals [35]. In non-invasive human studies, cross-sensory phase resetting of AC oscillations has also been reported, e.g., after seeing lip movements [33] or conversational hand gestures [36]. What would make cross-sensory phase resetting a particularly relevant modulatory mechanisms is that during normal communication the speakers facial and other gestured often slightly precede the onset of voice production.

BHFA signals are thought to reflect non-oscillatory broadband signals that correlate strongly with local neuronal firing [24-26], which offers a way to measure supra-threshold activation pattern in human SEEG [23] (for an alternative model, see, however, also [37]). Here, we mapped cross-sensory BHFA at the group level using a two-pronged strategy. First, BHFA was quantified from three anatomically normalized ROIs in each subject's surface-based cortical source estimates. Second, we analyzed BHFA in each subject's SEEG electrode contact that was closest to AC defined based on the HCP atlas. In the group analyses of source modeling data, traces of significant visually triggered BHFA signals, which were statistically significant in the FDR-adjusted estimates, were observed only in the area TPOJ. In contrast to the strong A and AV BHFA patterns, no significant visual BHFA effects were observed in AC, AAC, or the group analysis of the SEEG contact closest to AC. At the same time, in the exploratory, and statistically more liberal, uncorrected electrode-contact analysis in individual participants, significant BHFA to visual stimuli was observed in only one out of the 16 cases, in a participant whose sT contact closest to AC was, in fact, located in the upper bank of STS. These results are, overall, consistent with the observations in previous human ECoG [22] and SEEG studies [16] with smaller samples than in the present study of 16 participants. Taken together, these results suggest that at the population level, cross-sensory visual stimuli drive very weak, if any, suprathreshold neuronal activations in the vicinity of human AC.

The lack of cross-sensory BHFA in human AC is also in line with many previous studies that used more specific neurophysiological recordings in NPHs and other mammals. Laminar recordings suggested that cross-sensory somatosensory influences do not evoke increased MUA in the NHP primary AC [8]. Similar observation was reported in a subsequent NHP study where visual stimuli modulated single-unit activity (SUA) and MUA responses to auditory stimuli but did not trigger significantly increased firing when presented alone [6]. Opposite findings in NHP studies, which reported visually triggered firing effects in ACs have been attributed to byproducts of extensive training [17]. Beyond this, evidence for increases of suprathreshold activity by visual stimuli have been reported only in a small number of studies in other animal models [18, 20]. Our results are also broadly in line with previous human ECoG studies in the visual domain, which suggest very limited cross-sensory tactile BHFA effects even in higher-level visual areas of lateral occipital cortex [38]. All this is in line with the more conservative interpretation that cross-sensory stimuli play a modulatory rather than integrative role in human sensory cortices including ACs.

Potential routes for cross-sensory modulations and MSI effects in human AC have been previously proposed to include 1) feedback influences from higher-order attentional and/or multisensory convergence areas [33, 39], 2) direct inter-modal functional connectivity between low-level sensory areas [18, 40-42], and/or 3) subcortico-cortical feedback-type influences from non-specific thalamic nuclei to AC [43]. Many of the statistically strongest effects observed in the present study, including the beta-range ERDS and low-frequency MSI effects in AC, were found at relatively long latencies, that is, from more than 150-200 ms to up to 800 ms after the stimulus onset. Furthermore, in the case of low-frequency MSIs, the earliest signs of statistically significant effects occurred in TPOJ rather than in AC. A major portion of cross-sensory and MSI effects observed in the frequency domain could thus be explained by feedback from higher areas, instead of direct multisensory convergence. However, as estimated from the contact closest to AC, the cross-sensory ITPC effects were significant at the group level already ~ 100 ms after the stimulus onset. Furthermore, in our Laplacian-referenced LFP pattern analysis in the sT electrode, the cross-correlation analysis between LFP patterns to auditory vs. visual stimuli peaked at a lag of 55.2 ms, consistent with our previous MEG based estimate that cross-sensory visual responses follow the onset of auditory responses by about 55 ms [3]. These two last-mentioned observations leave room for an interpretation that more direct influences, based on direct cortico-cortical connectivity across different modalities or through subcortico-cortical influences via non-specific thalamic nuclei, play a role in cross-sensory modulation of human AC by simple stimuli. Naturally, what needs to be determined in future studies is to what extent this applies also to visual stimuli with strong auditory associations such as articulatory gestures.

In the classical sense, MSI refers to non-additive changes in the firing rates of individual neurons where the responses to bimodal audiovisual AV stimuli are stronger than the sum of the responses to auditory (A) and visual (V) stimuli alone [10]. In the present study, we however found evidence of significant MSI effects only in low-frequency oscillations. A potential complication for interpreting these MSI effects is that they primarily occurred at time-frequency windows in which we observed a similar ERDS pattern for all three stimulus types. It is thus likely that the significant effects in the AV vs. A+V contrast are not related to multisensory convergence, per se.

An important consideration in evaluating the present results is that subthreshold modulation effects including low-frequency LFP changes might be easier to detect than cross-sensory driving of neuronal activity. Here, the driving-type cross-sensory influences were, further, estimated using the BHFA, which is a correlate but not direct

measure of underlying neuronal activity [23-26] (see also [37]). However, it is worth noting that, as estimated by the group averages of baseline-normalized power to A stimuli, the effect sizes of individual subject's BHFA signals were at least as strong as, if not stronger, than effect sizes of lower-frequency LFPs that coincide the slow event-related responses observable in non-invasive EEG and MEG (**Fig. 4**).

In conclusion, we found direct intracranial neurophysiological evidence that cross-sensory visual stimuli modulate sub-threshold LFP signals and trigger low-frequency phase resetting in human ACs, with the cross-sensory LFP effects lagging those of unimodal A influences by about 50 ms. However, overall, in our large-sample study of intracranial human recordings, cross-sensory influences in ACs were somewhat weaker than we had anticipated. As determined using BHFA measures, there was little evidence of direct activation of AC by visual stimuli alone. Even when combined with concurrent A inputs, V stimuli drove only weak, if any, increases of BHFA effects in ACs themselves. Visually triggered BHFA influences were limited to the presumptively polymodal area TPOJ only. Finally, many of the most prominent cross-sensory modulatory influences, including the beta-range ERDS and concurrent MSI influences, occurred at relatively long latencies in ACs. Based on this, it seems that cross-sensory V inputs are involved in modulation of auditory responses, but no visual information processing, per se, takes place in human ACs. Analogously to the effects of selective attention, such cross-sensory modulatory influences could help suppress irrelevant sound features [44], enhance task-relevant rhythmic inputs [8, 45], and help guide attentional feedback resources to appropriate points of time to enhance auditory speech perception [46].

Methods

Participants: We studied sixteen patients (15–45 years, seven women) with pharmacologically intractable epilepsy who were undergoing clinically indicated intracranial SEEG recordings. All aspects such as implantation and positioning of electrodes, as well as durations of recordings, were based purely on clinical needs. All patients gave written informed consent before participating in this study. The study was approved by the Institute Review Board Taipei Veteran General Hospital. The details of participants' demographics and electrode implantation plans are provided in **Supplementary Table S1**. The sample size was selected to exceed the numbers of participants, which were studied in previous comparable intracranial human ECoG [22] and SEEG studies [16].

Experimental design: The participants were presented auditory (A), visual (V), or audiovisual (AV) stimuli in a randomized order, at a 1.2–2.8 s jittered inter-stimulus interval (ISI; mean 2 s). In an oddball design, the participant was asked to press a button with their right index finger upon detecting a target stimulus (p=0.1) among the non-target stimuli (**Fig. 1**). In A trials, the non-target stimulus was a 300-ms white-noise burst and the target stimulus an equally long 440 Hz pure tone. In V trials, the non-target stimuli consisted of static checkerboard stimuli (3.5 x 3.5 degrees of visual angle, 300 ms duration) and the target was the same checkerboard but overlapped with a central black diamond pattern. In AV trials, the participant heard and saw the combinations of A and V stimuli; the AV target was the combination of pure tone and checkerboard with a black diamond. The onset of the V stimulus preceded the A onset by 48 ms. Two six-minute runs of data were collected from each patient. In total, the subject was presented with 100 A, 100 V, and 100 AV stimuli. Auditory stimuli were delivered binaurally via insert earphones (Model S14, Sensimetrics, Gloucester, MA, USA) and the visual stimuli on a 17-inch computer screen (Asus MM17; ASUSTeK Computer Inc, Taipei, Taiwan) positioned at a viewing distance of 60 cm, controlled by using E-Prime (Psychology Software Tools, Sharpsburg, PA, USA).

Data acquisition: SEEG electrodes were placed solely based on the patient's clinical need, to identify epileptogenic zones. The patients were implanted with from 6 to 17 electrodes (0.3 mm diameter, 5-mm spacing; 6, 8, or 10 contacts per electrode Ad-Tech, Racine, WI, USA). SEEG data were collected at 2,048 samples/s. A white matter contact was used as a reference electrode in five of the 16 patients. In the remaining 11 patients, the scalp EEG electrode at the location FPz served as the initial reference [47-53]. However, in these 11 patients, the SEEG signals were re-referenced to a contact located in white matter before data analysis. **Figure 7a** depicts the distribution of all electrode contacts of all participants in a standard brain representation.

Two sets of anatomical MRIs were obtained, one before and another after each patient's electrode implantation surgery. We obtained pre-operative anatomical MRI images with a 3T MRI (Siemens Skyra, Siemens, Erlangen, Germany) using a MPRAGE sequence with TR/TI/TE/flip = 2530 ms/1100 ms/3.49 ms/7o, partition thickness = 1.33 mm, matrix = 256 x 256, 128 partitions, and FOV = 21 cm x 21 cm. The postoperative anatomical MRI was acquired at 1.5T (GE Signa HDxt system) with a fast spoiled gradient–recalled echo

sequence (TR/TE/TI 10.02/4.28/0 msec, flip angle 15°, matrix 256 * 256, bandwidth 31.2 kHz, view 256 × 256 mm, and axial slice thickness 1.0 mm).

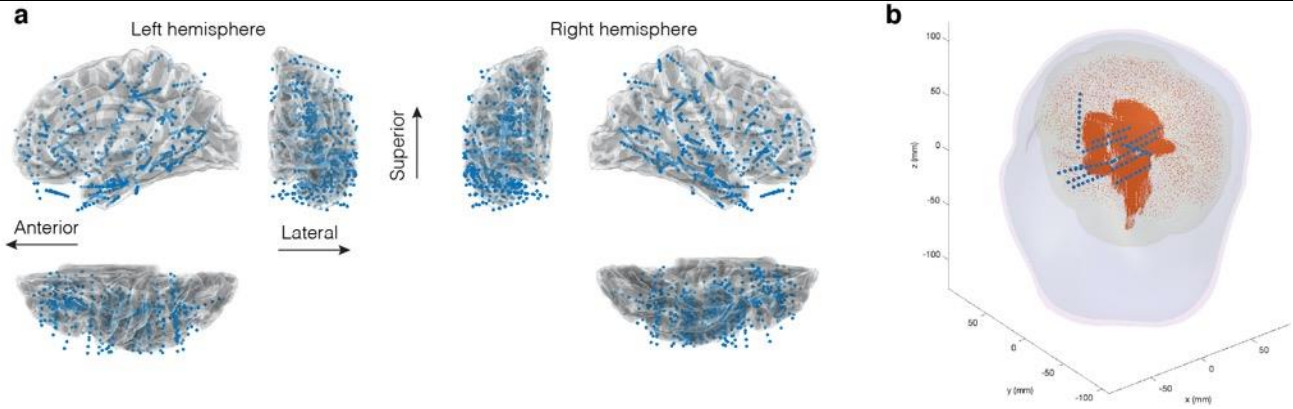


Figure 7. Electrode contact distribution and brain model diagram. (**a**) Locations of all electrode contacts, marked by the blue dots, are depicted in a template brain volume (MNI305). Lateral, frontal, and superior views of each hemisphere are shown. (**b**) A schematic of the source space (orange dots) including cortical and sub-cortical brain areas, locations of the electrode contacts (blue dots), and three anatomical boundaries (inner skull, outer skull, and outer scalp) for the lead field calculation using Boundary Element Model in a representative participant.

Preprocessing: SEEG preprocessing was performed by using the Matlab (MathWorks, Natick, MA) EEGLAB toolbox (version 14.1.0; http://sccn.ucsd.edu/eeglab/). Electrode contacts that were noisy or near epileptic lesion sites were excluded from the analyses based on initial visual screening. SEEG waveforms were notch filtered at 60 Hz to reduce power line artifacts and detrended when cropping the raw data into trials. Trials including epileptiform activity were excluded. SEEG recording was then segmented into epochs of 1500 ms duration with a 500-ms pre-stimulus baseline relative to the A, V, and AV stimulus onsets. SEEG epochs with deflections larger than six standard deviations from the epoch average were discarded. Finally, the evoked responses were averaged across epochs.

Source modeling: The coverage provided by depth electrode contacts is sparse and, due to clinical reasons, differs between patients. These two complications make it very difficult to analyze the extent of activations within any individual subjects and to conduct anatomically normalized group analyses, respectively. These problems could potentially be circumvented by intracranial source modeling of the SEEG data. Although SEEG source modeling has the same mathematical limitations associated with ill-posed inverse problems as MEG/EEG source

modeling, a crucial benefit is the directly measured depth information of the origins of the signals within the brain parenchyma. Another point is that the BHFA signal that is measurable with SEEG, and which is not visible to non-invasive MEG/EEG, has been reported to be highly focal [24, 26]. Therefore, in addition to conventional signal analyses, we employed inverse modeling of the intracranial source currents to facilitate anatomically normalized group analyses of cross-sensory influences on human AC.

SEEG source modeling was conducted using our recently published strategy [28]. The locations of electrode contacts in the individual's brain were identified from the post-surgery MRI, based on discrete dark image voxel clusters caused by the susceptibility artifact at each electrode contact. After specifying the distances between neighboring contacts and the number of contacts in an electrode, the electrode was manually aligned with the dark image voxel clusters in the post-surgery MRI using our in-house Matlab (Mathworks Inc., Natick MA, USA) software with a graphical user interface. Thereafter, contact locations were further optimized (within ±10 mm translation and ±2 degree rotation) by minimizing the sum of squares of image voxel values at all contact locations and their neighboring voxels within a 3x3x3 voxel cube in the post-surgery MRI using the Matlab *patternsearch* function. The SEEG electrode contact locations were then registered to the pre-surgery MRI, which was used to build Boundary Element Models (BEM's) required for the lead field calculation and to define locations of potential neural current sources. The inner-skull, outer-skull, and outer-scalp surfaces for the BEMs, as well as the cortical source spaces at the gray/white matter boundaries and sub-cortical source spaces, including thalamus, caudate, putamen, hippocampus, amygdala, and brain stem, were automatically reconstructed from the pre-surgery MRI by using FreeSurfer (http://surfer.nmr.mgh.harvard.edu). In each cortical source location, which were separated by about 5 mm from one another, we had three orthogonal neural current dipoles in $+x$, $+y$, and $+z$ directions. An example of the current source space electrode contacts and the inner-skull, outer-skull, and outer-scalp surfaces from a representative patient is shown in **Figure 7b**. The lead fields were calculated by using the *OpenMEEG* package (https://openmeeg.github.io/) [54, 55], with the relative conductivity values for air, scalp, brain parenchyma, and skull being 0, 1, 1, and 0.0125, respectively.

The measured sEEG data and the cortical current sources at time t are related to each other by $\mathbf{y}(t) = \mathbf{A}\,\mathbf{x}(t) + \mathbf{n}(t)$, where $\mathbf{y}(t)$ denotes the collection of sEEG data across electrode contacts, $\mathbf{x}(t)$ denotes the unknown current strength, and $\mathbf{n}(t)$ denotes noise. Electrode contacts potentially related to epileptic activity were excluded. The $\mathbf{x}(t)$ has 3 x m elements to describe the currents in three orthogonal directions at m brain locations

and **A** is the lead field matrix. For a unit current dipole source at location **r'** in the +*x*, +*y*, or +*z* direction, the electric potentials measured from all electrode contacts are reflected by **a**(**r'**) = [$\mathbf{a}_x$(**r'**), $\mathbf{a}_y$(**r'**), $\mathbf{a}_z$(**r'**)]. Assembling **a**(**r'**) across all possible current source locations created the lead field matrix **A**: **A** = [**a**($\mathbf{r'}_1$), **a**($\mathbf{r'}_2$), …, **a**($\mathbf{r'}_k$)], *k* =1, …., *d* where *d* denotes the total number of current dipole source locations.

To estimate **x**(*t*) using the minimum-norm estimate (MNE), we had $\boldsymbol{x}_{\boldsymbol{MNE}}(t) = \boldsymbol{RA}^T(\boldsymbol{ARA}^T + \lambda\boldsymbol{C})^{-1}\boldsymbol{y}(t)$, where **C** was the noise covariance matrix $\boldsymbol{C} = \langle \boldsymbol{n}(t)\boldsymbol{n}^{\boldsymbol{T}}(t)\rangle$**.** The operator ⟨·⟩ takes the ensemble average across the pre-stimulus interval (-200 ms to 0) with data concatenated across trials. The regularization λ tuned the balance between the strength of the estimated neuronal current strength and the discrepancy between the modeled and measured data. We chose $\lambda = 10$ [56].

Regions of Interest (ROI): Three ROIs were defined based on the Human Connectome Project (HCP) multimodal parcellation (MMP1) atlas, combined version [57], which was projected to the Freesurfer "fsaverage" standard brain surface representation. From the HCP atlas, we selected the ROIs "early auditory cortex" (AC), "auditory association cortex" (AAC), and "temporo-parieto-occipital junction" (TPOJ), which were subsequently resampled to each participant's individual cortical surface representation. For the group and individual-level TFR analyses, in each subject, we determined the contact that was closest to the HCP labels AC, AAC, and TPOJ. In the source modeling analysis of SEEG data, single-trial MNE time courses were calculated for the x, y, and z components at each dipole location and averaged across each ROI.

Time-frequency decomposition: Time-frequency representations (TFR) of SEEG power were calculated in Matlab by convolving individual trials with a dictionary of 7-cycle Morlet wavelets. In the case of ROIs of source estimates, the power TFRs were determined from the sums of squares of the amplitude values along three orthogonal dipole directions. For the group analyses, the power values were then averaged across trials and baseline normalized. In the supplementary individual-level analysis, we compared the power at each time/frequency element after the stimulus onset to the mean power during the respective pre-stimulus baseline period using paired t-tests. The inter-trial phase consistency (ITPC) across trials was calculated by dividing the mean amplitudes of the wavelet coefficients across the trials by the mean of absolute values of the amplitudes: $ITPC(f,t) = \frac{|\sum W(f,t)_m|}{\sum |W(f,t)_m|}$ where W denotes the wavelet coefficient, f is frequency, t is time, and m is the trial number.

Pattern analysis of superior temporal LFP data. We compared the similarity of LFP signals elicited by A and V stimuli by using 1,500 ms SEEG epochs, including a 500 ms pre-stimulus baseline period, averaged across trials from Laplacian-referenced contacts of the sT electrode. Only subjects with more than four adjacent contacts with acceptable signal quality in the sT electrode were included in this analysis (13 of the 16 participants). In each participant and task condition, Laplacian-referenced LFP data from m contacts and n time points were accumulated to two equally sized matrices. In each task condition, one matrix, $B_{m\times n}$, represented baseline time periods from -500 ms to -10 ms and the other matrix, $S_{m\times n}$, those from 10 ms to 500 ms relative to the stimulus onset. These condition and subject-specific matrices were reshaped to their respective vectors b_A, b_v, s_v, and s_A with $m\times n$ elements. Within each subject, Fisher-transformed Kendall correlation coefficients were then calculated across the six possible off-diagonal combinations of these four vectors (**Fig. 3b**). The Fisher-transformed correlation coefficients involving b_A or b_v were averaged in each subject and compared to the correlation between s_A and s_v using a paired t test. The purpose of this procedure was to mitigate spurious effects stemming from the similarity of noise levels in sT contacts within each subject. It is worth noting that all correlations across A and V trials reflect correspondences between fully independent samples.

Timing differences between LFPs in response to A vs. V stimuli were evaluated using a modified cross-correlation analysis. In each participant, we computed Kendall correlation coefficients between the LFP pattern generated for the A stimuli and increasingly later aspects of the LFP pattern generated for the V stimulus. To this end, analogously to the pattern analysis above, we first selected an auditory LFP matrix with m contacts and n time points representing the first 500 ms after the stimulus onset in the A condition, to be reshaped to a vector with $m\times n$ elements. In each iteration of the cross-correlation analysis, an equally sized $m\times n$ LFP matrix was then selected from the V condition, with the onset being displaced sample by sample from 0 to 500 ms, and reshaped to a vector. Kendall correlation coefficients between the vectors representing LFP patterns to A and V stimuli were then computed and Fisher transformed. The group analysis of statistical significance of the Fisher-transformed Kendall tau correlation coefficients was conducted using one-sample t-tests, corrected for multiple comparisons using the FDR procedure of [31]. In this analysis, we were specifically interested in finding the lag at which the group average of correlation coefficients reached its maximum. In both pattern correlation analyses, we converted the normalized correlation coefficients by calculating $\tau = \tanh(\bar{z})$ where $\bar{z}$ is the group average of

Fisher-transformed τ values in individual participants. The group-averaged τ were converted to Cohen d as described in [30].

Group analysis of TFRs. TFR values at each frequency were averaged across trials, baseline normalized, and taking the 10-based logarithm. ITPC values were first transformed to Z-score-like values by first calculating their p values [58]. Then, these p values were transformed to values of random variables following a standard normal distribution. We used a random-effect linear model to calculate the statistical inferences of TFR and ITPC in A, V, and AV across patients in both electrode and source analysis. A constant intercept was added to the random-effect linear modeling. We calculated the statistical significance for A, V, and AV conditions at each time-frequency grid. We also calculated the statistical significances between A, V, and AV conditions. All comparisons were corrected for the inflation of type-I error by controlling the false discovery rate (FDR).

Data for reference.

Codes used for this study are open to access at https://github.com/fahsuanlin/fhlin_toolbox/wiki.

Deidentified data utilized to yield our main results are available at https://dataverse.harvard.edu/privateurl.xhtml?token=d7216368-34bf-4a8a-858a-824b216385fb.

Any additional information required to reanalyze the data reported in this paper is available from the Corresponding Author.

Acknowledgements

Supported by: R01DC017991, R01DC016765, R01DC016915; Acad. Finland 276643, 298131, 308431. Russian Science Foundation 22-48-08002.

References

1. Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ. Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. Brain Res Cogn Brain Res. 2002;14(1):115-28. Epub 2002/06/14. doi: S0926641002000666 [pii]. PubMed PMID: 12063135.

2. Foxe JJ, Schroeder CE. The case for feedforward multisensory convergence during early cortical processing. Neuroreport. 2005;16(5):419-23. Epub 2005/03/17. doi: 00001756-200504040-00001. PubMed PMID: 15770144.

3. Raij T, Ahveninen J, Lin FH, Witzel T, Jääskeläinen IP, Letham B, et al. Onset timing of cross-sensory activations and multisensory interactions in auditory and visual sensory cortices. Eur J Neurosci. 2010;31(10):1772-82. Epub 2010/06/30. doi: 10.1111/j.1460-9568.2010.07213.x. PubMed PMID: 20584181; PubMed Central PMCID: PMC3008317.

4. Pekkola J, Ojanen V, Autti T, Jääskeläinen IP, Mottonen R, Tarkiainen A, et al. Primary auditory cortex activation by visual speech: an fMRI study at 3 T. Neuroreport. 2005;16(2):125-8. PubMed PMID: 15671860.

5. Schroeder CE, Foxe JJ. The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. Brain Res Cogn Brain Res. 2002;14(1):187-98. Epub 2002/06/14. doi: S0926641002000733 [pii]. PubMed PMID: 12063142.

6. Kayser C, Petkov CI, Logothetis NK. Visual modulation of neurons in auditory cortex. Cereb Cortex. 2008;18(7):1560-74. Epub 2008/01/09. doi: 10.1093/cercor/bhm187. PubMed PMID: 18180245.

7. Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK. Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. J Neurosci. 2005;25(20):5004-12. PubMed PMID: 15901781.

8. Lakatos P, Chen CM, O'Connell MN, Mills A, Schroeder CE. Neuronal oscillations and multisensory interaction in primary auditory cortex. Neuron. 2007;53(2):279-92. Epub 2007/01/17. doi: 10.1016/j.neuron.2006.12.011. PubMed PMID: 17224408; PubMed Central PMCID: PMCPMC3717319.

9. Wallace MT, Ramachandran R, Stein BE. A revised view of sensory cortical parcellation. Proc Natl Acad Sci U S A. 2004;101(7):2167-72. Epub 2004/02/10. doi: 10.1073/pnas.0305697101. PubMed PMID: 14766982; PubMed Central PMCID: PMCPMC357070.

10. Stein BE, Meredith MA. The merging of the senses. Cambridge: MIT Press; 1993.

11. Kayser C, Petkov CI, Augath M, Logothetis NK. Integration of touch and sound in auditory cortex. Neuron. 2005;48(2):373-84. Epub 2005/10/26. doi: 10.1016/j.neuron.2005.09.018. PubMed PMID: 16242415.

12. Konorski J. Integrative activity of the brain; an interdisciplinary approach. Chicago: University of Chicago Press; 1967.

13. Barlow HB. Single units and sensation: a neuron doctrine for perceptual psychology? Perception. 1972;1(4):371-94. Epub 1972/01/01. PubMed PMID: 4377168.

14. Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire PK, et al. Activation of auditory cortex during silent lipreading. Science. 1997;276(5312):593-6.

15. Foxe JJ, Morocz IA, Murray MM, Higgins BA, Javitt DC, Schroeder CE. Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. Brain Res Cogn Brain Res. 2000;10(1-2):77-83. Epub 2000/09/09. doi: 10.1016/s0926-6410(00)00024-0. PubMed PMID: 10978694.

16. Ferraro S, Van Ackeren MJ, Mai R, Tassi L, Cardinale F, Nigri A, et al. Stereotactic electroencephalography in humans reveals multisensory signal in early visual and auditory cortices. Cortex. 2020;126:253-64. Epub 2020/02/25. doi: 10.1016/j.cortex.2019.12.032. PubMed PMID: 32092494.

17. Brosch M, Selezneva E, Scheich H. Nonauditory events of a behavioral procedure activate auditory cortex of highly trained monkeys. J Neurosci. 2005;25(29):6797-806. Epub 2005/07/22. doi: 10.1523/JNEUROSCI.1571-05.2005. PubMed PMID: 16033889.

18. Bizley JK, Nodal FR, Bajo VM, Nelken I, King AJ. Physiological and anatomical evidence for multisensory interactions in auditory cortex. Cereb Cortex. 2007;17(9):2172-89. Epub 2006/12/01. doi: 10.1093/cercor/bhl128. PubMed PMID: 17135481.

19. Kobayasi KI, Suwa Y, Riquimaroux H. Audiovisual integration in the primary auditory cortex of an awake rodent. Neurosci Lett. 2013;534:24-9. Epub 2012/11/13. doi: 10.1016/j.neulet.2012.10.056. PubMed PMID: 23142716.

20. Morrill RJ, Hasenstaub AR. Visual Information Present in Infragranular Layers of Mouse Auditory Cortex. J Neurosci. 2018;38(11):2854-62. Epub 2018/02/15. doi: 10.1523/JNEUROSCI.3102-17.2018. PubMed PMID: 29440554; PubMed Central PMCID: PMCPMC5852663.

21. Giard MH, Peronnet F. Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. J Cogn Neurosci. 1999;11(5):473-90. Epub 1999/10/08. PubMed PMID: 10511637.

22. Mercier MR, Molholm S, Fiebelkorn IC, Butler JS, Schwartz TH, Foxe JJ. Neuro-oscillatory phase alignment drives speeded multisensory response times: an electro-corticographic investigation. J Neurosci. 2015;35(22):8546-57. Epub 2015/06/05. doi: 10.1523/JNEUROSCI.4527-14.2015. PubMed PMID: 26041921.

23. Parvizi J, Kastner S. Promises and limitations of human intracranial electroencephalography. Nat Neurosci. 2018;21(4):474-83. Epub 2018/03/07. doi: 10.1038/s41593-018-0108-2. PubMed PMID: 29507407.

24. Manning JR, Jacobs J, Fried I, Kahana MJ. Broadband shifts in local field potential power spectra are correlated with single-neuron spiking in humans. J Neurosci. 2009;29(43):13613-20. Epub 2009/10/30. doi: 10.1523/JNEUROSCI.2041-09.2009. PubMed PMID: 19864573; PubMed Central PMCID: PMCPMC3001247.

25. Miller KJ. Broadband spectral change: evidence for a macroscale correlate of population firing rate? J Neurosci. 2010;30(19):6477-9. Epub 2010/05/14. doi: 10.1523/JNEUROSCI.6401-09.2010. PubMed PMID: 20463210.

26. Ray S, Maunsell JH. Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. PLoS Biol. 2011;9(4):e1000610. Epub 2011/05/03. doi: 10.1371/journal.pbio.1000610. PubMed PMID: 21532743; PubMed Central PMCID: PMCPMC3075230.

27. Besle J, Fischer C, Bidet-Caulet A, Lecaignard F, Bertrand O, Giard MH. Visual activation and audiovisual interactions in the auditory cortex during speech perception: intracranial recordings in humans. J Neurosci.

2008;28(52):14301-10. Epub 2008/12/26. doi: 10.1523/JNEUROSCI.2875-08.2008. PubMed PMID: 19109511.

28. Lin F-H, Lee H-J, Ahveninen J, Jääskeläinen IP, Yu H-Y, Lee C-C, et al. Distributed source modeling of intracranial stereoelectroencephalographic measurements. Neuroimage. 2021;230:117746. Epub 2021/01/18. doi: 10.1016/j.neuroimage.2021.117746 PubMed PMID: 33454414; PubMed Central PMCID: PMCPMC8044004.

29. Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science. 2001;293(5539):2425-30. Epub 2001/09/29. doi: 10.1126/science.1063736. PubMed PMID: 11577229.

30. Walker DA. Converting Kendall's tau for correlational or meta-analytic analyses. Journal of Modern Applied Statistical Methods. 2003;2(2):525-30. doi: 10.22237/jmasm/1067646360.

31. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. 1995;57:289-300.

32. Stanford TR, Stein BE. Superadditivity in multisensory integration: putting the computation in context. Neuroreport. 2007;18(8):787-92. Epub 2007/05/02. doi: 10.1097/WNR.0b013e3280c1e315. PubMed PMID: 17471067.

33. Luo H, Liu Z, Poeppel D. Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. PLoS Biol. 2010;8(8):e1000445. Epub 2010/08/17. doi: 10.1371/journal.pbio.1000445. PubMed PMID: 20711473; PubMed Central PMCID: PMCPMC2919416.

34. Atilgan H, Town SM, Wood KC, Jones GP, Maddox RK, Lee AKC, et al. Integration of Visual Information in Auditory Cortex Promotes Auditory Scene Analysis through Multisensory Binding. Neuron. 2018;97(3):640-55 e4. Epub 2018/02/06. doi: 10.1016/j.neuron.2017.12.034. PubMed PMID: 29395914; PubMed Central PMCID: PMCPMC5814679.

35. Megevand P, Mercier MR, Groppe DM, Zion Golumbic E, Mesgarani N, Beauchamp MS, et al. Crossmodal Phase Reset and Evoked Responses Provide Complementary Mechanisms for the Influence of Visual Speech in Auditory Cortex. J Neurosci. 2020;40(44):8530-42. Epub 2020/10/08. doi: 10.1523/JNEUROSCI.0555-20.2020. PubMed PMID: 33023923; PubMed Central PMCID: PMCPMC7605423.

36. Biau E, Torralba M, Fuentemilla L, de Diego Balaguer R, Soto-Faraco S. Speaker's hand gestures modulate speech perception through phase resetting of ongoing neural oscillations. Cortex. 2015;68:76-85. Epub 2015/01/18. doi: 10.1016/j.cortex.2014.11.018. PubMed PMID: 25595613.

37. Leszczynski M, Barczak A, Kajikawa Y, Ulbert I, Falchier AY, Tal I, et al. Dissociation of broadband high-frequency activity and neuronal firing in the neocortex. Sci Adv. 2020;6(33):eabb0977. Epub 2020/08/28. doi: 10.1126/sciadv.abb0977. PubMed PMID: 32851172; PubMed Central PMCID: PMCPMC7423365.

38. Quinn BT, Carlson C, Doyle W, Cash SS, Devinsky O, Spence C, et al. Intracranial cortical responses during visual-tactile integration in humans. J Neurosci. 2014;34(1):171-81. Epub 2014/01/02. doi: 10.1523/JNEUROSCI.0532-13.2014. PubMed PMID: 24381279; PubMed Central PMCID: PMCPMC3866483.

39. Smiley JF, Hackett TA, Ulbert I, Karmas G, Lakatos P, Javitt DC, et al. Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. J Comp Neurol. 2007;502(6):894-923. Epub 2007/04/21. doi: 10.1002/cne.21325. PubMed PMID: 17447261.

40. Budinger E, Heil P, Hess A, Scheich H. Multisensory processing via early cortical stages: Connections of the primary auditory cortical field with other sensory systems. Neuroscience. 2006;143(4):1065-83. Epub 2006/10/10. doi: 10.1016/j.neuroscience.2006.08.035. PubMed PMID: 17027173.

41. Rockland KS, Ojima H. Multisensory convergence in calcarine visual areas in macaque monkey. Int J Psychophysiol. 2003;50(1-2):19-26. Epub 2003/09/27. doi: S0167876003001211 [pii]. PubMed PMID: 14511833.

42. Falchier A, Schroeder CE, Hackett TA, Lakatos P, Nascimento-Silva S, Ulbert I, et al. Projection from visual areas V2 and prostriata to caudal auditory cortex in the monkey. Cereb Cortex. 2010;20(7):1529-38. Epub 2009/10/31. doi: 10.1093/cercor/bhp213. PubMed PMID: 19875677; PubMed Central PMCID: PMCPMC2882821.

43. Hackett TA, De La Mothe LA, Ulbert I, Karmos G, Smiley J, Schroeder CE. Multisensory convergence in auditory cortex, II. Thalamocortical connections of the caudal superior temporal plane. J Comp Neurol. 2007;502(6):924-52. Epub 2007/04/21. doi: 10.1002/cne.21326. PubMed PMID: 17444488.

44. Kauramaki J, Jääskeläinen IP, Hari R, Mottonen R, Rauschecker JP, Sams M. Lipreading and covert speech production similarly modulate human auditory-cortex responses to pure tones. J Neurosci.

2010;30(4):1314-21. Epub 2010/01/29. doi: 10.1523/JNEUROSCI.1950-09.2010. PubMed PMID: 20107058; PubMed Central PMCID: PMCPMC2832801.

45. Schroeder CE, Lakatos P, Kajikawa Y, Partan S, Puce A. Neuronal oscillations and visual amplification of speech. Trends Cogn Sci. 2008;12(3):106-13. Epub 2008/02/19. doi: 10.1016/j.tics.2008.01.002. PubMed PMID: 18280772; PubMed Central PMCID: PMCPMC3987824.

46. Zion Golumbic E, Cogan GB, Schroeder CE, Poeppel D. Visual input enhances selective speech envelope tracking in auditory cortex at a "cocktail party". J Neurosci. 2013;33(4):1417-26. Epub 2013/01/25. doi: 10.1523/JNEUROSCI.3675-12.2013. PubMed PMID: 23345218; PubMed Central PMCID: PMCPMC3711546.

47. Cam SL, Ranta R, Caune V, Korats G, Koessler L, Maillard L, et al. SEEG dipole source localization based on an empirical Bayesian approach taking into account forward model uncertainties. Neuroimage. 2017;153:1-15. Epub 2017/03/23. doi: 10.1016/j.neuroimage.2017.03.030. PubMed PMID: 28323161.

48. Caune V, Ranta R, Le Cam S, Hofmanis J, Maillard L, Koessler L, et al. Evaluating dipolar source localization feasibility from intracerebral SEEG recordings. Neuroimage. 2014;98:118-33. Epub 2014/05/06. doi: 10.1016/j.neuroimage.2014.04.058. PubMed PMID: 24795155.

49. Jonas J, Frismand S, Vignal JP, Colnat-Coulbois S, Koessler L, Vespignani H, et al. Right hemispheric dominance of visual phenomena evoked by intracerebral stimulation of the human visual cortex. Hum Brain Mapp. 2014;35(7):3360-71. Epub 2014/04/16. doi: 10.1002/hbm.22407. PubMed PMID: 24733699; PubMed Central PMCID: PMCPMC6869193.

50. Jonas J, Jacques C, Liu-Shuang J, Brissart H, Colnat-Coulbois S, Maillard L, et al. A face-selective ventral occipito-temporal map of the human brain with intracerebral potentials. Proc Natl Acad Sci U S A. 2016;113(28):E4088-97. Epub 2016/06/30. doi: 10.1073/pnas.1522033113. PubMed PMID: 27354526; PubMed Central PMCID: PMCPMC4948344.

51. Koessler L, Cecchin T, Colnat-Coulbois S, Vignal JP, Jonas J, Vespignani H, et al. Catching the invisible: mesial temporal source contribution to simultaneous EEG and SEEG recordings. Brain Topogr. 2015;28(1):5-20. Epub 2014/11/30. doi: 10.1007/s10548-014-0417-z. PubMed PMID: 25432598.

52. Mittal S, Barkmeier D, Hua J, Pai DS, Fuerst D, Basha M, et al. Intracranial EEG analysis in tumor-related epilepsy: Evidence of distant epileptic abnormalities. Clin Neurophysiol. 2016;127(1):238-44. Epub 2015/10/24. doi: 10.1016/j.clinph.2015.06.028. PubMed PMID: 26493495.

53. Rikir E, Koessler L, Gavaret M, Bartolomei F, Colnat-Coulbois S, Vignal JP, et al. Electrical source imaging in cortical malformation-related epilepsy: a prospective EEG-SEEG concordance study. Epilepsia. 2014;55(6):918-32. Epub 2014/04/08. doi: 10.1111/epi.12591. PubMed PMID: 24702598.

54. Gramfort A, Papadopoulo T, Olivi E, Clerc M. OpenMEEG: opensource software for quasistatic bioelectromagnetics. Biomed Eng Online. 2010;9:45. Epub 2010/09/08. doi: 10.1186/1475-925X-9-45. PubMed PMID: 20819204; PubMed Central PMCID: PMCPMC2949879.

55. Kybic J, Clerc M, Abboud T, Faugeras O, Keriven R, Papadopoulo T. A common formalism for the integral formulations of the forward EEG problem. IEEE Trans Med Imaging. 2005;24(1):12-28. Epub 2005/01/11. doi: 10.1109/tmi.2004.837363. PubMed PMID: 15638183.

56. Lin FH, Belliveau JW, Dale AM, Hämäläinen MS. Distributed current estimates using cortical orientation constraints. Hum Brain Mapp. 2006;27(1):1-13. PubMed PMID: 16082624.

57. Glasser MF, Coalson TS, Robinson EC, Hacker CD, Harwell J, Yacoub E, et al. A multi-modal parcellation of human cerebral cortex. Nature. 2016;536(7615):171-8. Epub 2016/07/21. doi: 10.1038/nature18933. PubMed PMID: 27437579; PubMed Central PMCID: PMCPMC4990127.

58. Fisher NI. Statistical Analysis of Circular Data. Cambridge, UK: University Press; 1993.

Supplementary Information

Subject	Gender	Age	Seizure onset zones	N of electrodes	Contacts included in analysis
1	M	17	Right OFC and MTL	13	83
2	M	25	Left MTL	12	71
3	F	29	Left MTL	14	103
4	F	21	Left MTL	12	70
5	M	25	Right parietal lobe	11	63
6	F	45	Right MTL	10	58
7	M	18	Right MTL	10	68
8	F	39	Bilateral MTL	10	70
9	F	16	Right MTL	11	69
10	M	15	Right parietal lobe	6	38
11	M	15	Right parietal lobe	16	86
12	M	38	Right MTL	11	72
13	M	23	Right MTL	13	94
14	F	33	Right MTL	14	57
15	F	23	Left MTL and insula	11	68
16	M	34	Right occipital lobe	17	98

Supplementary Table S1. Demographic information for the 16 patients included in analysis. **Abbreviations**: M, male; F, female; OFC, orbitofrontal cortex; MTL, medial temporal lobe.

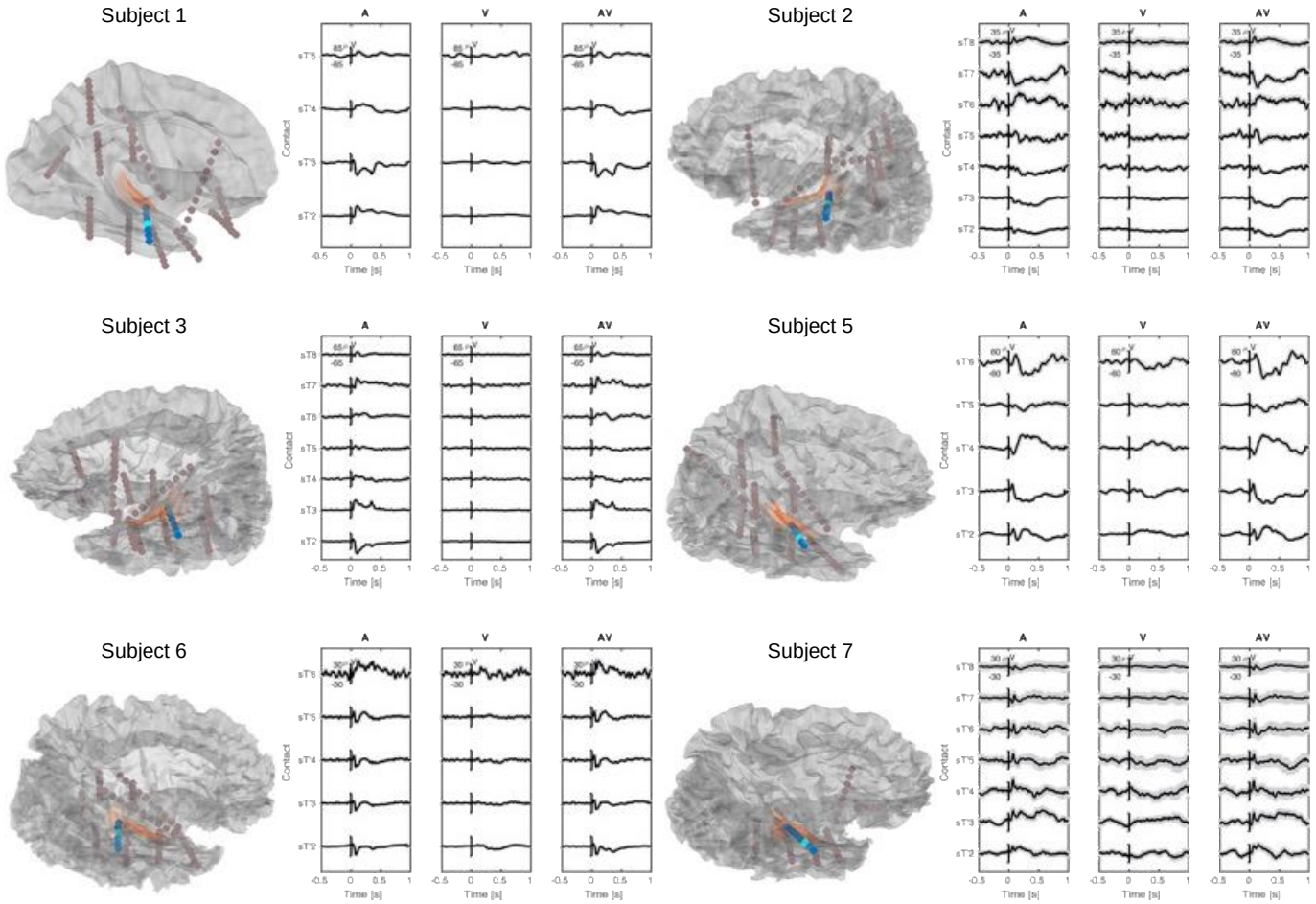


Figure 1. Intracranial local field potential (LFP) data data in Subjects 1, 2, 3, 5, 6, and 7. The locations of the superior temporal (sT) electrode contacts are shown on the left. Trial-averaged response waveforms are shown. The gray shading shows the standard error of mean (SEM) across the trials. The data show a bipolar derivation with each contact being referenced to its nearest medial neighbor.

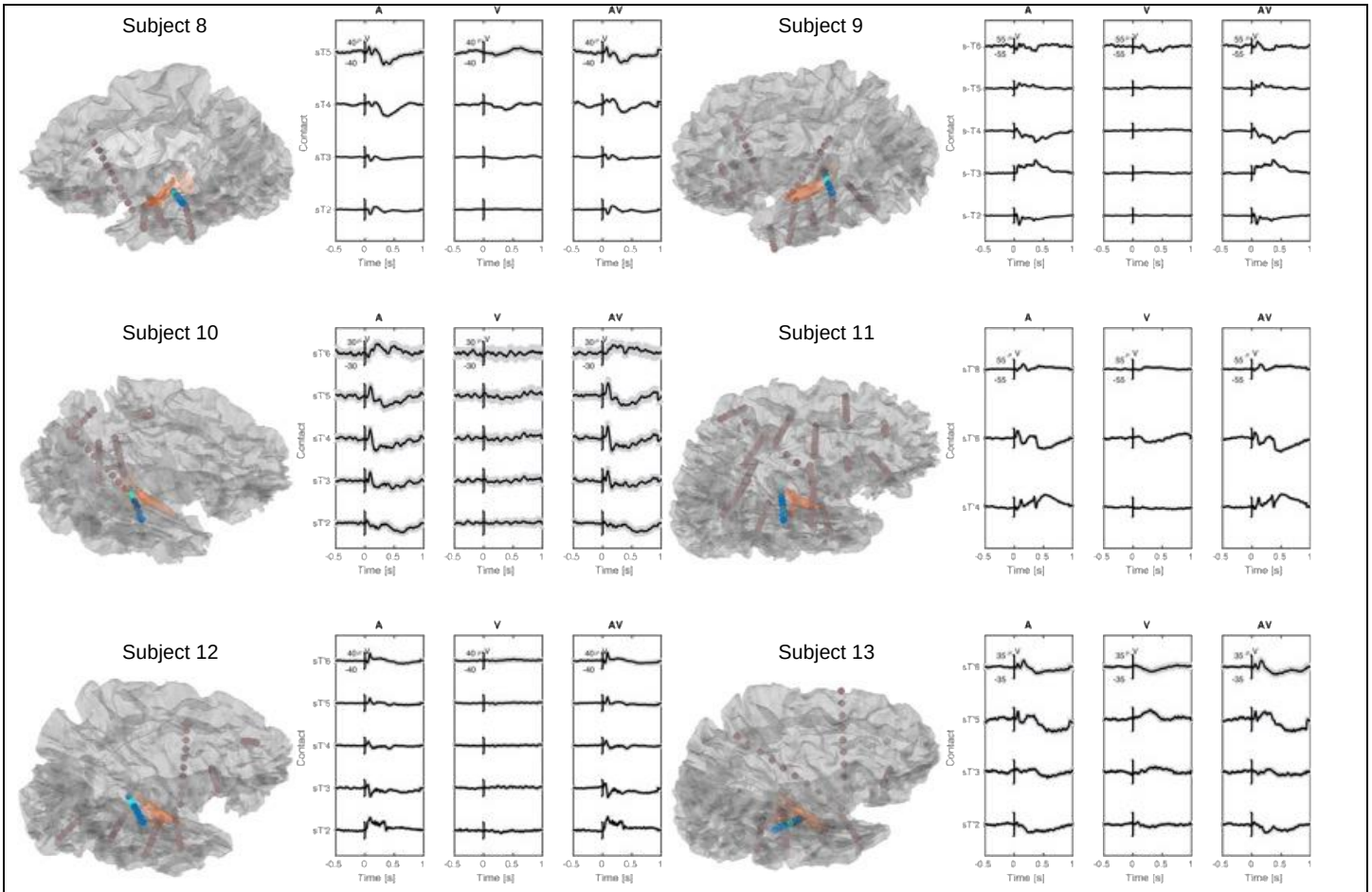


Figure S2. Intracranial local field potential (LFP) data data in Subjects 8-13. The locations of the superior temporal (sT) electrode contacts are shown on the left. Trial-averaged response waveforms are shown. The gray shading shows the standard error of mean (SEM) across the trials. The data show a bipolar derivation with each contact being referenced to its nearest medial neighbor.

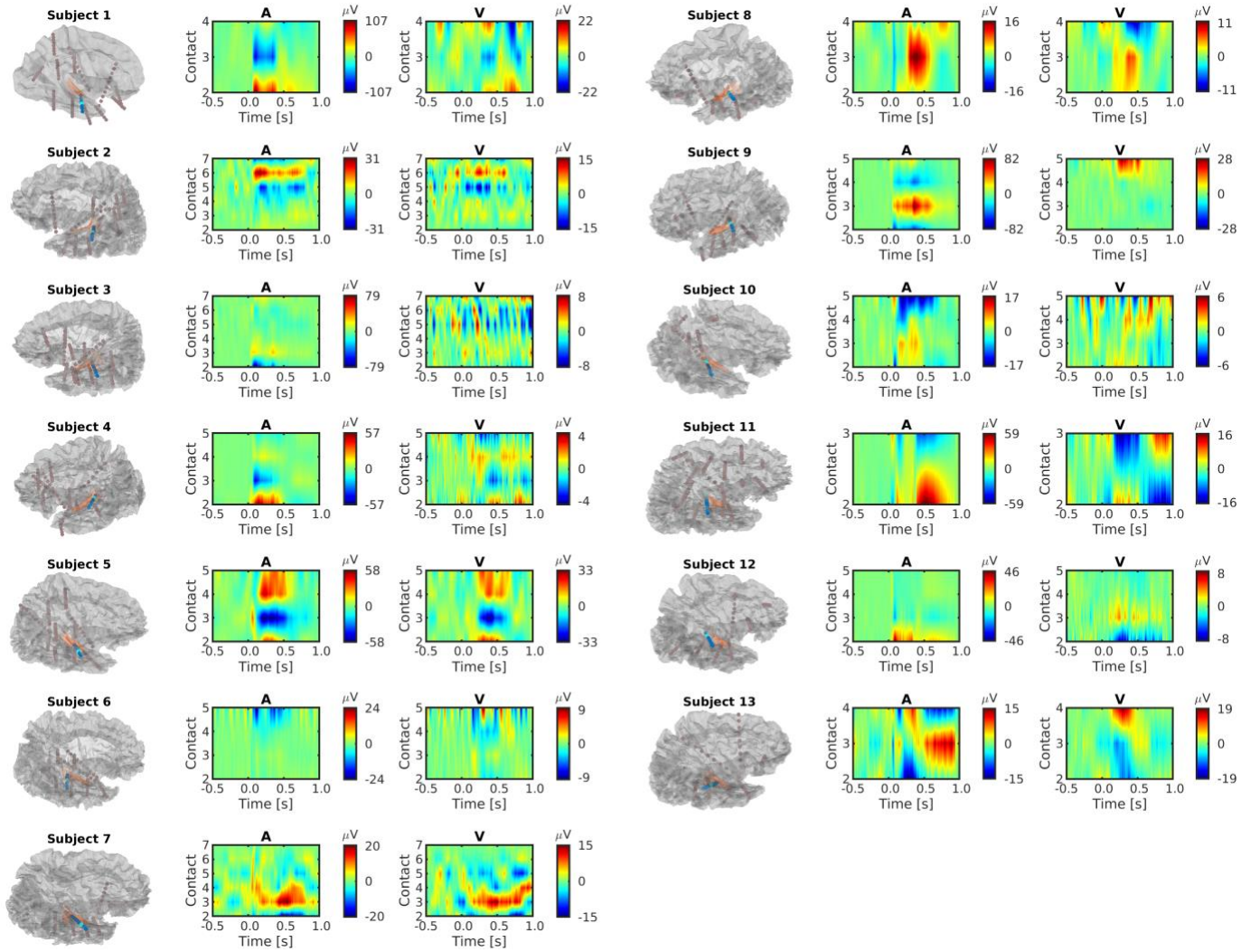
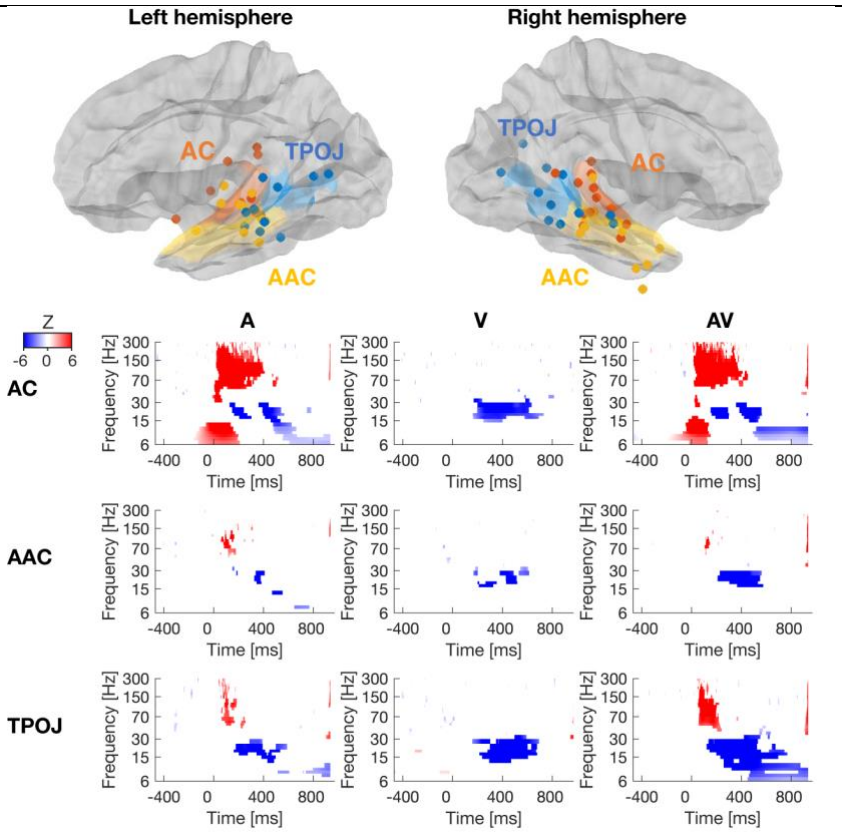
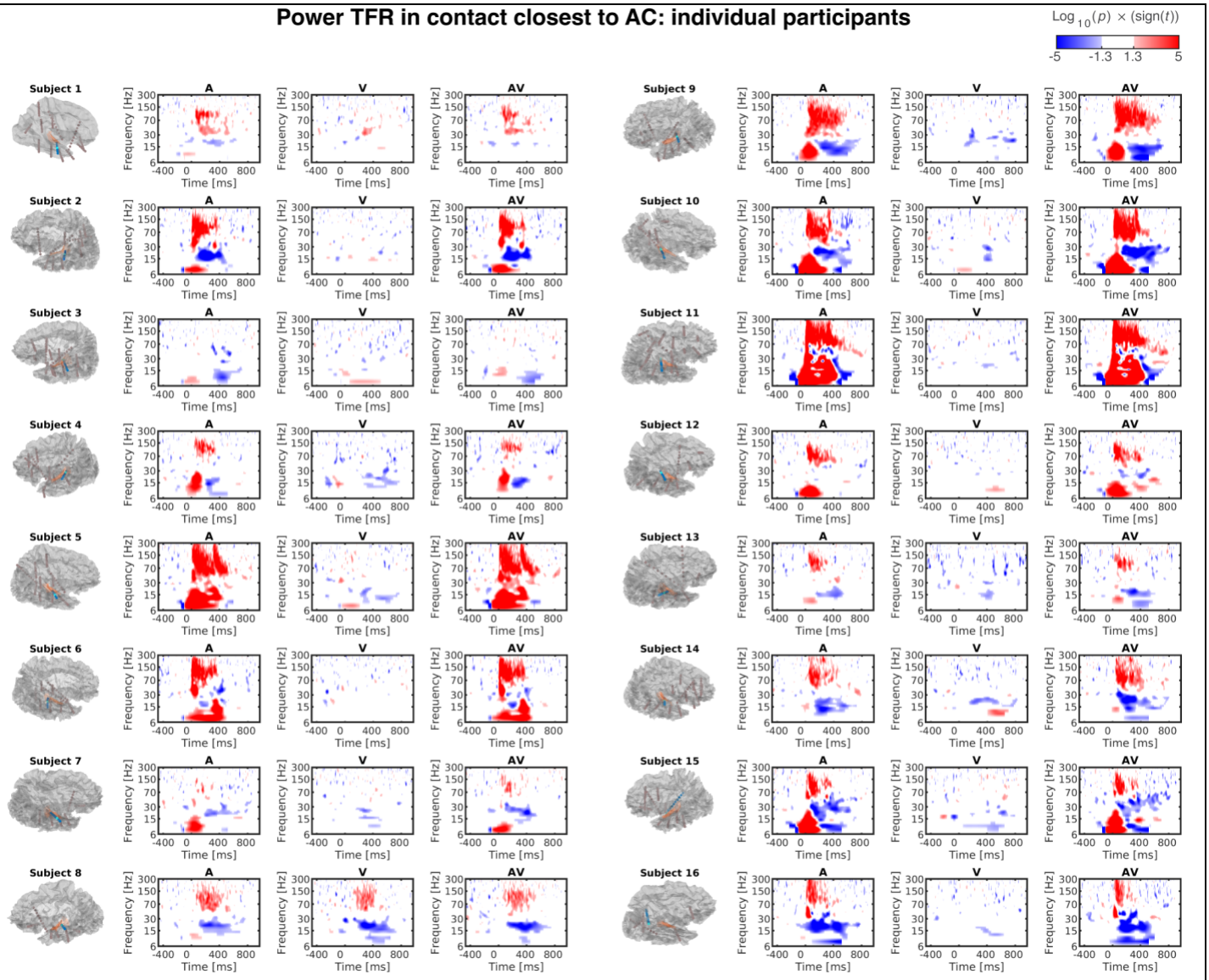


Figure S3. Laplacian-referenced responses analyzed across the contacts of a superior temporal SEEG electrode to A and V stimuli individual participants. The data of 13 participants (out of total cohort of 16) who had at least four adjacent contacts in the superior temporal sT electrode were included in this analysis. For each participant, contacts of sT electrode are marked with blue. Notably, there are several cases with remarkable similarity in the LFP pattern to A and V stimuli. Related to **Fig. 1** of the main manuscript.

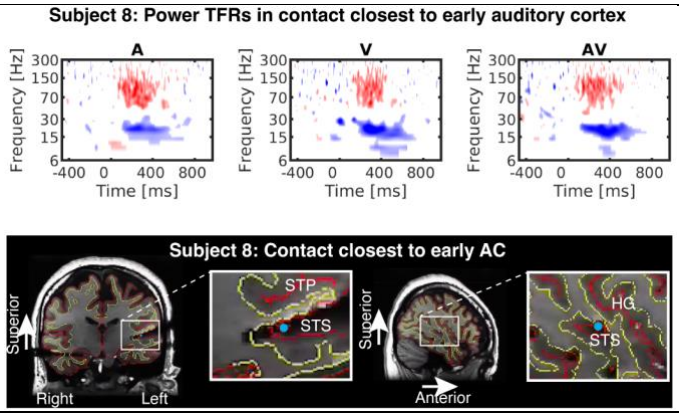


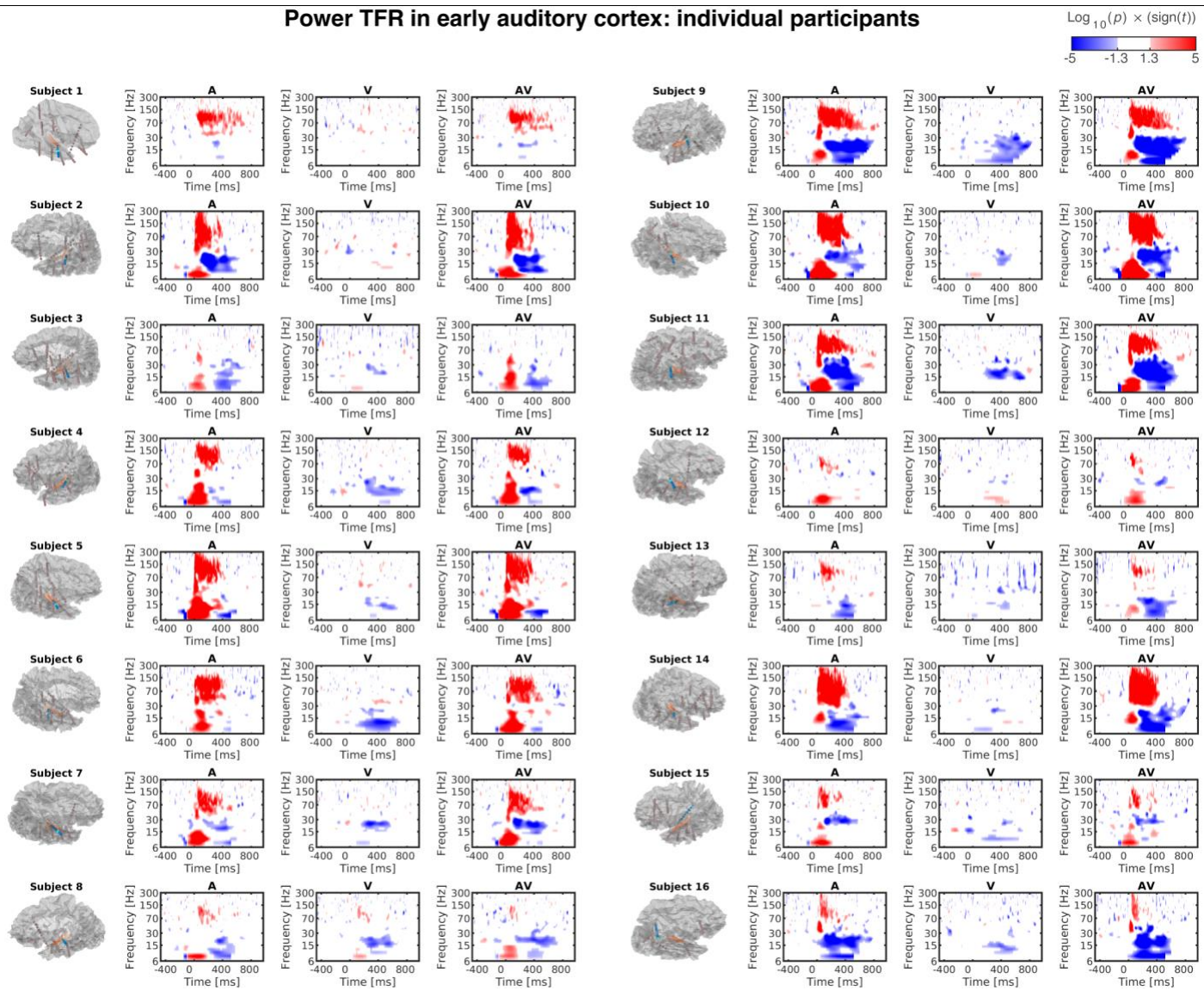
Suppl. Fig. **S4**. Group analysis of power TFRs in electrode contacts. Each subject's contacts closest to early auditory cortex (AC), auditory association cortex (AAC), and temporo-parieto-occipital junction (TPOJ) are marked with orange, yellow, and blue dots, respectively. Significant BHF (>70 Hz, FDR-adjusted p<0.05) emerges to auditory (A) and audiovisual (AV) stimuli only. Cross-sensory V stimuli suppress beta (15-30 Hz) power 200 ms after the stimulus onset in all studied contact sets, consistent with the ROI analysis (**Fig. 2**).



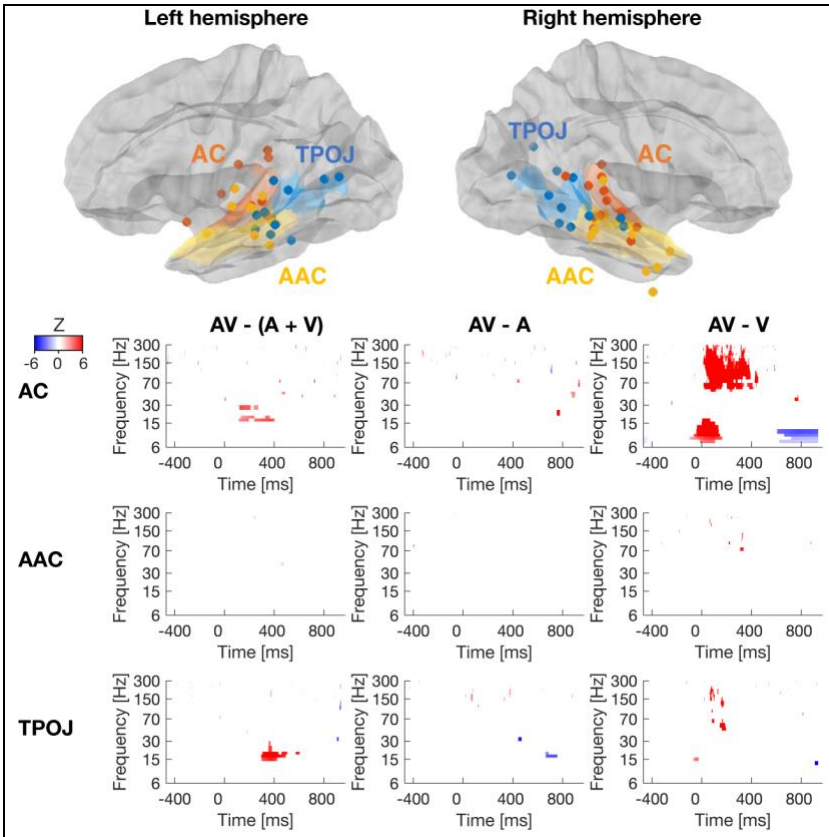
Suppl. Fig. S5. Power TFRs in the electrode contact closest to AC in individual participants. The statistical significance of a t test relative to the pre-stimulus baseline is shown, thresholded at the uncorrected p<0.05.

Suppl. Fig. S6. Power TFR and the location of contact closest to AC in Subject 8. Significant (uncorrected p<0.05) BHF activity was observed only one out of the 16 subjects. In this subject, the contact closest to AC was in STS (cyan dot in the MRI views), a putative multisensory processing area.

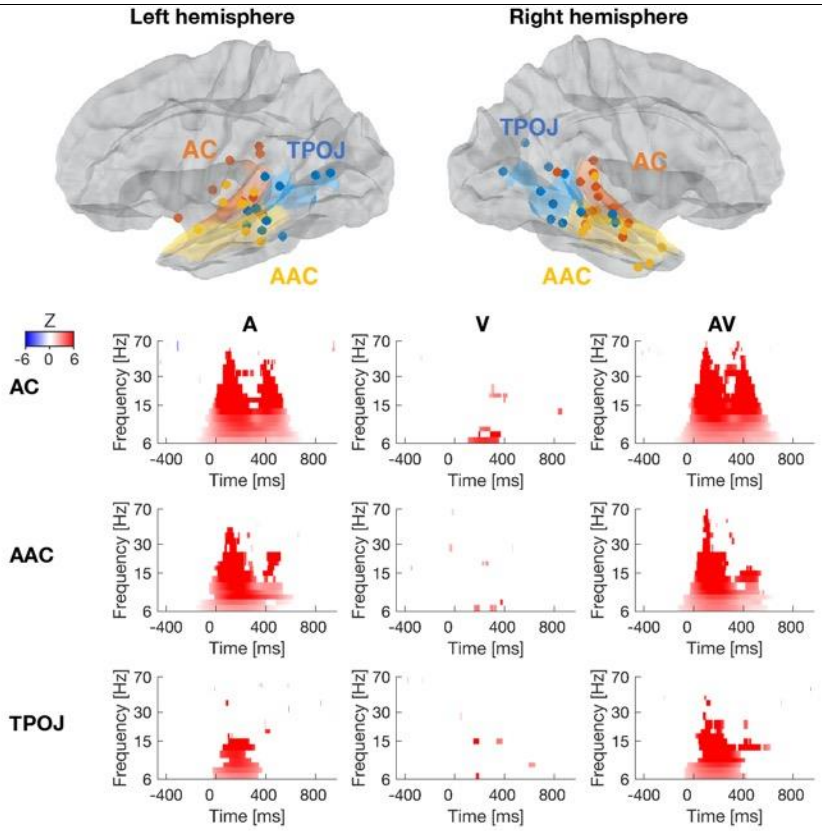




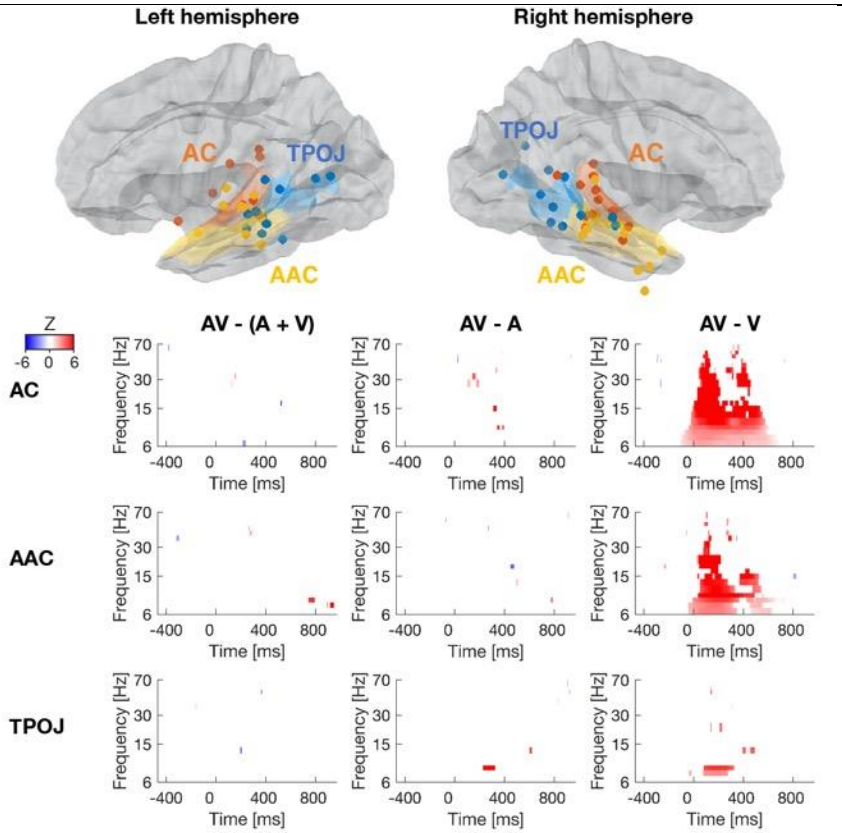
Suppl. Fig. S7. Power TFRs in AC source-modeling ROI in individual participants. The statistical significance of a t test relative to the pre-stimulus baseline is shown, thresholded at the uncorrected $p<0.05$.



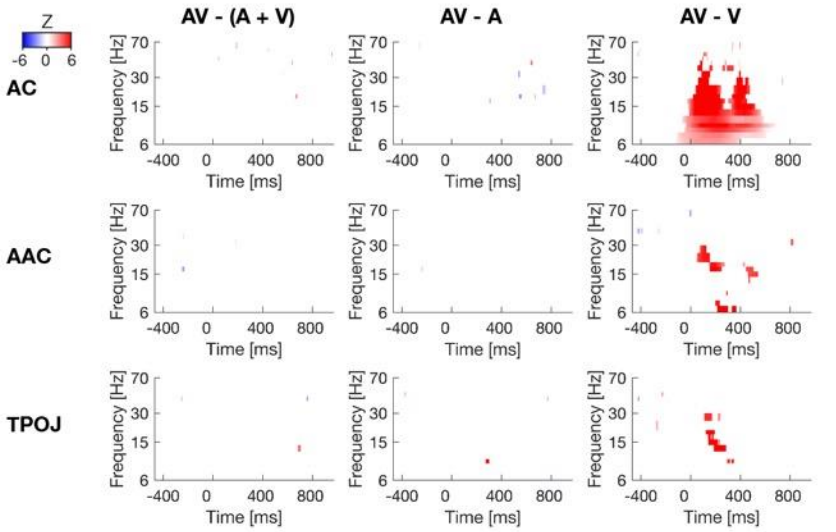
Suppl. Fig. **S8**. Group analysis of power TFR contrasts between conditions in electrode contacts. Each subject's contacts closest to AC are marked with orange dots, contacts closest to AAC with yellow dots, and contacts closest to TPOJ with blue dots. Indices of statistically significant superadditive multisensory interactions occur at the beta range in the AV-(A+V) contrast in contacts closest to AC and TPOJ (FDR-adjusted $p<0.05$). The effects in AC occur almost 200 ms earlier than those in TPOJ. The contrast AV-V shows the anticipated enhancement of BHF and early-latency theta/alpha in AC. Overall, the effects beyond AC are somewhat weaker than those in the corresponding ROI analysis **(Fig. 3**).



Suppl. Fig. **S9**. Group analysis of ITPC in electrode contacts. Each subject's contacts closest to AC are marked with orange dots, contacts closest to AAC with yellow dots, and contacts closest to TPOJ with blue dots. In addition to the strong ITPC effects to A and AV stimuli, there is also evidence of significant increase of low-frequency ITPC to V stimuli (FDR-adjusted $p<0.05$). The effects in contacts closest to AAC and TPOJ are somewhat weaker than those in the corresponding source-modeling ROI analysis (**Fig. 4**).



Suppl. Fig. **S10**. Group analysis of ITPC contrasts across conditions in electrode contacts. Each subject's contacts closest to AC are marked with orange dots, contacts closest to AAC with yellow dots, and contacts closest to TPOJ with blue dots. Excluding the anticipated differences in the AV-V contrast, only weak evidence of group differences in AV-(A+V) and AV-A contrasts was found.



Suppl. Fig. **S11**. Group analysis of ITPC contrasts across conditions in source-modeling ROIs. Excluding the anticipated differences in the AV-V contrast, only weak evidence of group differences in AV-(A+V) and AV-A contrasts was found in this analysis.