

Primary Generators of Visually Evoked Field Potentials Recorded in the Macaque Auditory Cortex

Yoshinao Kajikawa,^{1,2} John F. Smiley,³ and Charles E. Schroeder^{1,2}

¹Translational Neuroscience Division, Nathan S. Kline Institute for Psychiatric Research, Orangeburg, New York 10962, ²Cognitive Science and Neuromodulation Program, Department of Neurosurgery, Columbia University, New York, New York 10032, and ³Department of Neurochemistry, Nathan S. Kline Institute for Psychiatric Research, Orangeburg, New York 10962

Prior studies have reported “local” field potential (LFP) responses to faces in the macaque auditory cortex and have suggested that such face-LFPs may be substrates of audiovisual integration. However, although field potentials (FPs) may reflect the synaptic currents of neurons near the recording electrode, due to the use of a distant reference electrode, they often reflect those of synaptic activity occurring in distant sites as well. Thus, FP recordings within a given brain region (e.g., auditory cortex) may be “contaminated” by activity generated elsewhere in the brain. To determine whether face responses are indeed generated within macaque auditory cortex, we recorded FPs and concomitant multiunit activity with linear array multielectrodes across auditory cortex in three macaques (one female), and applied current source density (CSD) analysis to the laminar FP profile. CSD analysis revealed no appreciable local generator contribution to the visual FP in auditory cortex, although we did note an increase in the amplitude of visual FP with cortical depth, suggesting that their generators are located below auditory cortex. In the underlying inferotemporal cortex, we found polarity inversions of the main visual FP components accompanied by robust CSD responses and large-amplitude multiunit activity. These results indicate that face-evoked FP responses in auditory cortex are not generated locally but are volume-conducted from other face-responsive regions. In broader terms, our results underscore the caution that, unless far-field contamination is removed, LFPs in general may reflect such “far-field” activity, in addition to, or in absence of, local synaptic responses.

Key words: auditory cortex; face response; field potential; inferotemporal cortex; macaque monkey; volume conduction

Significance Statement

Field potentials (FPs) can index neuronal population activity that is not evident in action potentials. However, due to volume conduction, FPs may reflect activity in distant neurons superimposed upon that of neurons close to the recording electrode. This is problematic as the default assumption is that FPs originate from local activity, and thus are termed “local” (LFP). We examine this general problem in the context of previously reported face-evoked FPs in macaque auditory cortex. Our findings suggest that face-FPs are indeed generated in the underlying inferotemporal cortex and volume-conducted to the auditory cortex. The note of caution raised by these findings is of particular importance for studies that seek to assign FP/LFP recordings to specific cortical layers.

Introduction

Field potentials (FPs) reflect neuronal ensemble activity (Schroeder et al., 1998; Kajikawa and Schroeder, 2011; Buzsáki et al., 2012), and often this activity is subthreshold to, or otherwise not

evident in, action potentials (APs). In a large, primarily Ohmic medium, such as the brain, FPs can be approximated by a spatial integration of synaptically mediated transmembrane currents that are weighted by their spatial proximity to a measuring point (Kajikawa and Schroeder, 2015). Ordinarily, this would mean that, activity of neurons near the recording electrode (near field) is better represented in the FP than that of distant neurons. However, remote (far-field) activity can also influence FP, especially when it is stronger than local activity. When there are multiple neuronal populations that differ in their temporal activity patterns, FPs at sites in between exhibit temporal patterns that are mixtures of contributions from in those populations (Kajikawa and Schroeder, 2015).

Received Dec. 12, 2016; revised July 21, 2017; accepted July 25, 2017.

Author contributions: Y.K. and C.E.S. designed research; Y.K. performed research; Y.K. contributed unpublished reagents/analytic tools; Y.K. and J.F.S. analyzed data; Y.K., J.F.S., and C.E.S. wrote the paper.

This study was supported by National Institutes of Health Grants R01DC015780 and R01DC011490. We thank Dr. Mark Klinger for veterinary assistance; and Dr. Deborah Ross for technical support.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Yoshinao Kajikawa, Nathan Kline Institute, 140 Old Orangeburg Road, Orangeburg, NY 10962. E-mail: ykajikawa@nki.rfmh.org.

DOI:10.1523/JNEUROSCI.3800-16.2017

Copyright © 2017 the authors 0270-6474/17/3710139-15\$15.00/0

Visual activation/modulation of low level auditory cortex is considered to be an early-stage substrate of audiovisual (AV) integration (Ghazanfar and Schroeder, 2006; Driver and Noesselt, 2008); and in macaques, it has been shown by several techniques measuring cortical activity: single/multiunit activity (MUA), local field potential (LFP), and functional imaging. Studies report no significant change in neuronal firing rate after visual stimuli, but rather, speeding of auditory response onsets (Chandrasekaran et al., 2013), and/or increase in the information carried by auditory-evoked firing patterns (Kayser et al., 2010). In contrast, visual stimuli, particularly faces, not only modulate auditory LFP responses (Ghazanfar et al., 2005), but also evoke LFPs by themselves (Kayser et al., 2007a, 2008; Hoffman et al., 2008). Visual-evoked LFPs with little to no change in local neuronal firing suggest that either (1) visual stimuli alone evoke subthreshold synaptic responses or (2) visual-evoked LFP reflect far-field, rather than local activity.

To better understand the nature of visual responses in auditory cortex, we recorded laminar profiles of FPs and concomitant MUA from auditory cortex in macaques performing auditory and visual tasks. Because previous studies argued that conspecific faces were the most effective modulators of responses to conspecific vocal sounds in auditory cortex (Ghazanfar et al., 2005; Hoffman et al., 2008), the present study focused on the responses to macaque monkey faces. FP recordings were augmented by current source density (CSD) analysis. Because it eliminates effects of volume conduction, CSD is a better indicator of local activity than FPs alone (Kajikawa and Schroeder, 2011). Whereas FP responded strongly to faces, associated CSD and MUA responses were negligible, indicating little to no local contribution to generation of the face-evoked visual FP response. Instead, visual FP responses grew larger with depth within and below auditory cortex. Tracking the FP responses below auditory cortex revealed visual MUA and CSD responses in the inferotemporal (IT) cortex. Our results indicate that face-evoked FP responses in auditory cortex are primarily far-field reflections of responses generated in IT. Thus, while FP methods clearly provide unique and valuable information on population neuronal activity, strict localization of their sources requires spatial differentiation over 2 or more recordings at millimeter/submillimeter scales.

Materials and Methods

All procedures were approved by the Institutional Animal Care and Use Committee of the Nathan Kline Institute.

Subjects. Three macaque monkeys (*Macaca mulatta*; Monkey P: female; Monkey G and Monkey W: males) were implanted with headposts and recording chambers (one per hemisphere, both sides in Monkey P and right side in Monkeys G and W) using aseptic surgical techniques. Based on presurgical MRI, the chambers were positioned to aim penetrations perpendicular to auditory cortices on the lower bank of the lateral sulcus.

Behavioral paradigms. Monkeys were trained to perform the auditory and visual oddball tasks in a sound-attenuated chamber. The monkey started a trial by pulling a lever that brought up a gray rectangle area (17.8×11.4 degrees) on the screen (see Fig. 1A). The monkey then maintained gaze position within the window for at least 400 ms to initiate a sequence of sensory events that started with a static image appearing in the window for 900 ms, followed by a 500 ms nontarget stimulus. The sequence of the static image and the nontarget stimulus repeated randomly 3–6 times in each trial; then a target stimulus appeared. Stimuli could be auditory (A)-alone, visual (V)-alone, or bimodal (AV) versions of conspecific vocalizations (for details, see Stimuli). The modality and exemplar of stimuli differed randomly across trials but were constant within each trial. After the first nontarget, the duration of intervals presenting the static image between the offset of the prior stimulus and the

onset of the following stimulus were randomly chosen from 600, 750, 900, 1050, and 1200 ms to reduce potential effects of cognitive entrainment (Lakatos et al., 2009). Monkeys were required to maintain gaze position within a window until the target stimulus was presented regardless of the stimulus modality, and responded to the target manually by releasing the lever to obtain an aqueous reward. Every response was followed by a >1 s blank period. Gaze position was monitored continuously using Eyelink-1000 (SR-Research). Stimulus deliveries, tracking of the lever, and reward deliveries were controlled using the Experiment Builder (SR-Research).

Stimuli. Movie clips of 9 exemplars of macaque vocalizations (courtesy of Prof. Romanski, University of Rochester, and others recorded at the Nathan Kline Institute) were used. Clips were trimmed to start from the onset of vocalizing face movements and to last for 500 ms using the Adobe Premiere (Adobe Systems), and separated into a visual track (15 frames, 29.97 fps) and an audio track (44.1 kHz sample rate) using the utility software of the Experiment Builder (SR-Research). Visual tracks were zoomed in using Adobe Premiere, so that faces occupied a central zone of ~ 10 degrees diameter circle area when displayed on screen, and further edited to blacken the backgrounds behind the monkeys' head and below the collar using a custom script in MATLAB (The MathWorks).

For the AV trials, a movie track and an audio track were presented synchronously as nontarget and target stimuli. Sounds in the audio tracks started at 131–257 ms later than the onset of the visual tracks. In the following figures in Results, the timing of auditory stimuli and cortical responses was shown with the same delayed timing as experiments. The visual-auditory offset is due to the delay from the initial visible articulation gesture to the onset of accompanying vocal sounds (Schroeder et al., 2008; Chandrasekaran et al., 2011). For V-alone trials, a movie track was presented without audio track. For A-alone trials, an audio track was presented concurrently with a movie track of a static image (see below) that was shown from the period just before the stimulus. In effect, no noticeable change occurred in the image during the time the sounds were played. The delayed onset timing of sound relative to the null-motion movie track was the same as stimuli in AV trials. For both the AV and V-alone trials, the repetition of a movie track was interleaved with a static image. For the A-alone trials, the static image remained still on the screen throughout repetitions of an audio track.

There were two task variants that differed in the static image that appeared before the first stimulus and between stimuli within a trial (see Fig. 1B). In Task 1, the static image was a scramble of nonblack pixels in the image of the first frame of the following movie track. In Task 2, the static image was an image of the first frame of the following movie track. The use of these static images created difference in the timing of the abrupt face onsets in a trial between Task 1 and 2. In Task 1, faces appeared abruptly at every onset of nontarget and target stimuli. In Task 2, abrupt appearance of faces occurred only at the beginning of trials, and movie tracks started with a smooth transition from preexisting static face to move without interruption.

All visual stimuli were presented on a monitor (FlexScan F930, Eizo), 90 cm in front of the monkey. Images and movies were presented within a rectangular window (17.8×11.4 degrees) at the center of blank screen. Auditory stimuli were delivered from either loudspeakers (Tannoy Precision 6P) placed on both sides of the monitor through an amplifier (Ashly ne800) for Monkey P and W, or the magnetic speakers (FF1, Tucker Davies Technologies) placed at 4 inches from ears for Monkey G.

Electrophysiological recordings. Recordings were conducted during task performance. Single or dual electrodes comprised of a linear array of 23 electrical channels spaced at either 100 or 200 μm apart (0.3–0.5 M Ω at 1.0 kHz) were used (U-Probe, Plexon). Each channel recorded FPs (0.1–500 Hz) and MUA (200–5000 Hz, further bandpass, zero phase shift digital filtering 300–1000 Hz, 48 dB/octave and rectifying) simultaneously. The resultant signals were sampled at 2 kHz. Even though these MUA signals do not isolate individual spikes, their magnitude reflects the frequency of neuronal firing, due to high-frequency characteristics of APs (Legatt et al., 1980; Kayser et al., 2007b). A metal pin immersed in the saline filling the recording chamber was used as a reference electrode.

The CSD was calculated from FPs recorded from three adjacent electrode channels by numerical differentiation to approximate the second-

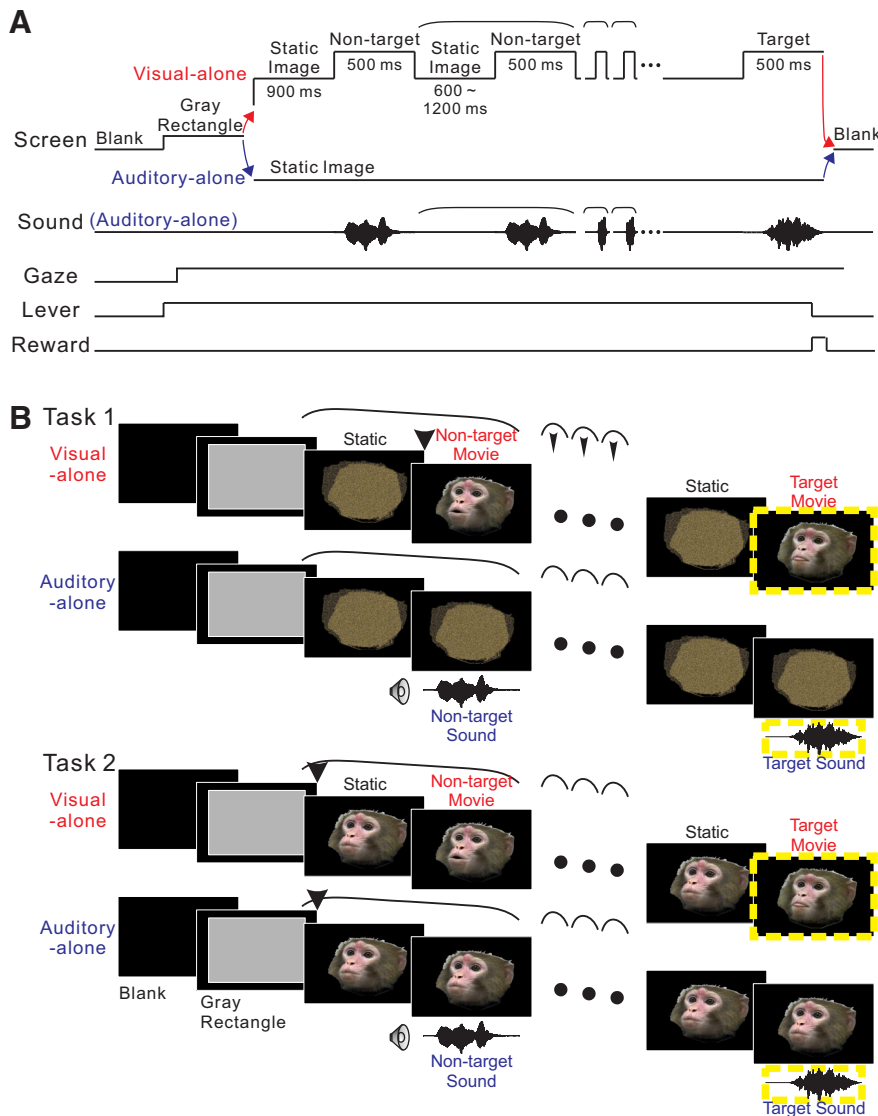


Figure 1. The tasks and stimuli. **A**, Behavioral paradigm. Each trial was initiated when the monkey pulled a lever, bringing up a rectangular gray window. The monkey had to keep its gaze within the window for 400 ms to start a trial. Trials started with a static image followed by nontarget stimuli, which could be just a vocal sound (auditory or A-alone), just a movie clip (visual or V-alone), or bimodal (AV). For A-alone trials, the static image remained on the screen until the end of each trial. After repeating the sequence of static image–nontarget stimuli randomly from 3 to 6 times, we presented an oddball (target) stimulus that differed in its sound or movie content from the nontarget (**B**, circumscribed by yellow dashed lines). In the figure, the third and fourth stimuli are temporally shrunken. Monkeys had to release the lever upon detection of the target to receive a reward. This was followed by a >1000 ms break period, before another trial could be initiated. **B**, Stimuli during Tasks 1 and 2. While both tasks used same nontarget stimulus and repeated a sequence of static image and a nontarget stimulus, they differed in the static image that appeared before the first and between following nontargets. The image in Task 2 was the first frame of the following movie clip, whereas in Task 1 the image was the scramble of the first frame. Consequently, a face abruptly appeared at the onset of every movie clip in a trial of Task 1, but in Task 2 a face appeared abruptly only once before the first movie clip. Arrowheads indicate the timing when such abrupt onset of face could occur in both Tasks. In an A-alone trial, the static image was taken from the first clip of the movie clip that was a visual counterpart of the following vocal sound.

order spatial derivative (Schroeder et al., 1998). Whereas spiking in MUA can reflect activity in fibers of passage at a recording site, CSD analysis provides a more definitive assessment of a truly local synaptic response (Kajikawa and Schroeder, 2011). CSD response occurs whether or not the summation of synaptic activity results in APs in the local neuronal population (Schroeder et al., 1998).

Linear array electrodes were inserted in parallel to one another through a grid of guide tubes. In the recordings from auditory cortex using arrays of 23 channels evenly spaced over 2.2 mm, CSD components of sensory responses can be flanked by flat or radically weak CSD, indicating that arrays crossed all 6 layers of an active cortex. Given the di-

mensions of arrays (2.2 mm), crossing all cortical layers suggests that penetration angles were nearly orthogonal to the layers.

In each experiment, after positioning the electrode arrays, but before engaging the monkey in the experimental task, we examined stimulus preferences at each recording site using pure tones (353.55 Hz to 32 kHz with 0.5 octave intervals) and broad-band noise (BBN) delivered in a quasi-random order at 60 dB SPL (duration: 100 ms, stimulus onset asynchrony: 625 ms, averaging 50–100 responses for each tone and BBN stimulus). The best frequency of MUA responses to the set of tones was identified for each recording site (Kajikawa and Schroeder, 2011) and used to derive the tonotopic map. The positions of recording sites ranged from AP0 to AP16 across animals and were mostly in the core (primary auditory [A1], and rostral [R]) or caudal belt areas (caudal-medial [CM], and caudal-lateral [CL]). Gradients of best frequency were used to define the borders between areas R and A1, and between areas A1 and CM or CL (Recanzone et al., 2000). Transitions to broad (vs sharp) tuning profile, along with preference for BBN over tones were used to define the border between A1 and the belt areas.

Penetrations that tracked FPs below auditory cortex found depths of several millimeters corresponding to the white matter where temporal fluctuations of FP occurred uniformly with no CSD correlates, although high-amplitude MUA was often noted. High-amplitude MUA in the absence of concomitant CSD is considered a signature of white matter (Schroeder et al., 1998). Below the white matter, inversions of FP with concomitant CSD components with MUA appeared. Based on its position, this expanse corresponded to the upper and lower banks of the STS, with the quiet region in between corresponding to the STS itself. The upper and lower banks of STS are termed areas TPO and TEa, respectively (Baylis et al., 1987). MUA, FP, and CSD in the lower bank of the STS revealed consistent face responses across penetrations: this location was identified as part of the IT cortex. The AP position of recordings in IT ranged between AP1 and AP14.

Auditory cortical layers were identified based on the responses to pure tones or BBN using standard criteria as described previously (Kajikawa and Schroeder, 2011). Layer 4 of auditory cortex is marked by the presence of a brief, short latency current sink with a concomitant increase in multiunit activity (Stein-schneider et al., 1992), as the earliest sinks of sensory responses are typically found at middle cortical layers in sensory cortices (Jellema et al., 2004; McLaughlin and Juliano 2005; Müller-Preuss and Mitzdorf 1984; Takeuchi et al., 2011; van Kerkoerle et al., 2017). Based on known anatomy (Hackett et al., 2001, 2014), the positions of the supragranular and infragranular layers can be identified with respect to the position of layer 4. Similar spatiotemporal progression of signals occurs in other sensory cortices (Schroeder et al., 1998; Lipton et al., 2010) and was observed in IT. For both responses to vocal sounds in the auditory cortex and responses to faces in IT, the supragranular response corresponds to the most superficial source/sink pair, and this current flow configuration generates the most

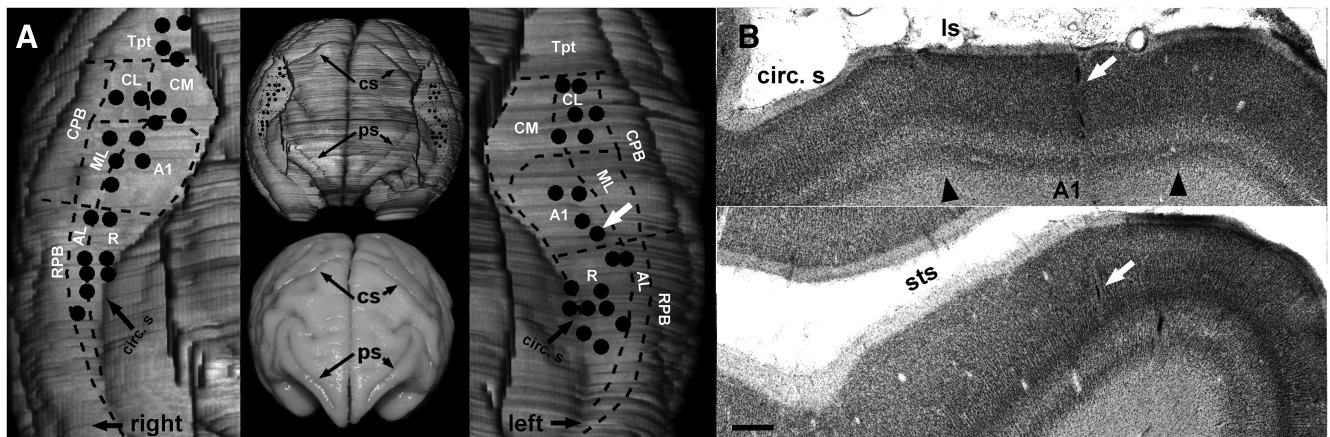


Figure 2. Location of electrode penetrations in Monkey P. **A**, Reconstruction of Monkey P's brain, in which the surface of the superior temporal gyrus was exposed by cutting away the parietal and frontal opercula. Left, Right, Auditory areas at higher magnification. Dashed lines indicate borders of auditory areas that were identified in Nissl-stained and parvalbumin-immunolabeled sections. Black dots indicate the electrode penetration sites, identified in Nissl-stained and immunolabeled sections. White arrow indicates the penetration into auditory belt that is shown in **B**. **B**, Nissl-stained sections of auditory cortex (top) and the lower bank of the STS (bottom), showing the electrode penetration (white arrows) indicated by white arrow in **A**. Black arrowheads indicate the borders of auditory core; RPB, rostral parabelt area; sts, superior temporal sulcus; Tpt, temporal parietotemporal area. Scale bar: **B**, 0.5 mm.

prominent FP inversions. At depths a few hundred micrometers deeper than the supragranular CSD sink, smaller CSD sink and concurrent MUA responses start before the onset of the supragranular sink/source pair. The depths of earlier CSD sink were identified as the granular layer responses. In auditory cortex, penetrations depicting the progression of best frequency with depth, suggestive of highly oblique penetration angles, were excluded from analyses. The angles as well as positions of recording tracks were confirmed by postmortem reconstruction (see Histology).

Data analysis. Analyses were conducted for auditory responses to nontarget stimuli during A-alone trials in Task 1 (averaging 60–200 responses for each stimulus) and Task 2 (30–120 responses for each stimulus), and the visual responses to nontarget stimuli during V-alone trials in Task 1 (60–200 responses for each stimulus) or the onsets of static face images in the beginning of all trials in Task 2 (60–300 responses for each stimulus). Responses to stimuli were calculated by averaging epochs of all stimulus presentations. For the FPs, signals were digitally bandpass filtered offline using a third-order Butterworth filter at cutoff frequencies of 1 and 256 Hz. Peak amplitudes were estimated after subtracting the mean amplitude of a 100 ms baseline period prior and common to all peaks. For nonlocal responses, such as visual responses in auditory cortex and auditory responses in IT, the peak latency of FP components was measured in layer 4. For local responses, such as auditory responses in auditory cortex and visual responses in IT, timing shifts occur within those cortices due to volume conduction (Kajikawa and Schroeder, 2015). Thus, peak latencies were derived at channels just below auditory cortex or just above IT where no timing shifts occur. Magnitude of MUA response was quantified as the average MUA elevation during a 10 ms period around the peak from baseline level (average of 50 ms before the face onset) at channels within the granular layer.

To summarize the spatial distributions of peak amplitudes across penetration sites, depth-amplitude profiles of FP peaks were first normalized by the square root of the mean of the absolute amplitude values across all 23 channels within each site. Depths between sites were aligned relative to the depth of inversion of a peak as described. For MUA responses, the amplitudes of MUA were around the peak timing of granular layer MUA, normalized by the largest amplitude across depth for each site, and aligned relative to the inversion depth between sites. Confidence intervals (CIs) of normalized amplitude distributions were derived by a bootstrap procedure (1000 resampling).

Volume conductor model. CSD and FP signals at all channels were bandpass filtered with cutoff frequencies at 1.5 and 128 Hz, then subtracted the mean of 100 ms baseline. Both FP and CSD responses were epoched from 50 to 400 ms after the onset of stimuli. Visual responses usually delay by >50 ms. Whereas auditory responses delay by <50 ms,

the onset of sound was delayed by 120–200 ms from the stimulus onset. Thus, the same epoch included visual responses to face or auditory responses to sound. The volume conductor model was applied to the spatiotemporal patterns of CSD epochs to reproduce the patterns of the observed FP epochs (Kajikawa and Schroeder, 2015). The model was formulated for FP at each depth, d_k , of a k -th channel as follows:

$$FP(k, t) = A \sum_j \frac{CSD(j, t)}{\sqrt{(r_h \cdot d)^2 + |d_j - d_k|^2}}$$

in which $d = d_{k+1} - d_k$ is the distance between neighboring contacts of array electrode, and r_h represents the ratio of the horizontal spread of CSD from array electrode to d and is the sole free parameter that maximizes the similarity score. The similarity score was derived as the Frobenius inner product of the spatiotemporal profiles of the model-derived (predicted) FP responses and that of the recorded (observed) FP responses. The profiles were normalized by the square root of mean squared values, for which averaging was done across both time and depth dimensions as $FP(k, t) / \sqrt{|FP(k, t)|^2}$. The similarity score as the products of those normalized profiles takes value from 1 for the profiles of same shapes to -1 for the profiles of reversed shapes, regardless of the magnitude of $FP(k, t)$ or the scale factor A in $FP(k, t)$.

Statistical tests. Nonparametric statistics tests were used to compare response parameters between groups and conditions. Kruskal–Wallis test for independent samples was used. $p = 0.05$ was considered as the criterion level in all tests. When a test was significant, Tukey's HSD test of ranks was used for multiple comparisons.

Histology. Monkey P was deeply anesthetized by a lethal dose (100 mg/kg) of sodium pentobarbital and perfused intracardially with saline (room temperature), followed by 4 L of ice-cold 4% paraformaldehyde in 0.1 M phosphate buffer. The brain was removed, and placed in buffered 20% glycerol for 1 week, before sectioning at 80 μ m thickness. The brain frontal and temporal poles were frozen to a sliding microtome, so that the plane of section was approximately perpendicular to the axis of the lateral sulcus (Lipton et al., 2010). A digital camera mounted over the microtome recorded the appearance and location of every section, and these images were used for 3D reconstruction using the Volume Viewer plugin in ImageJ software (see Fig. 2A). Series of every 12th section were processed for Nissl staining or parvalbumin immunolabeling. Electrode penetration sites were identified microscopically in the histological sections (see Fig. 2B), and their locations registered on the block-face digital images and 3D reconstructions. Individual sites were identified by comparison of their distribution with the penetration coordinates in the electrode guide matrices.

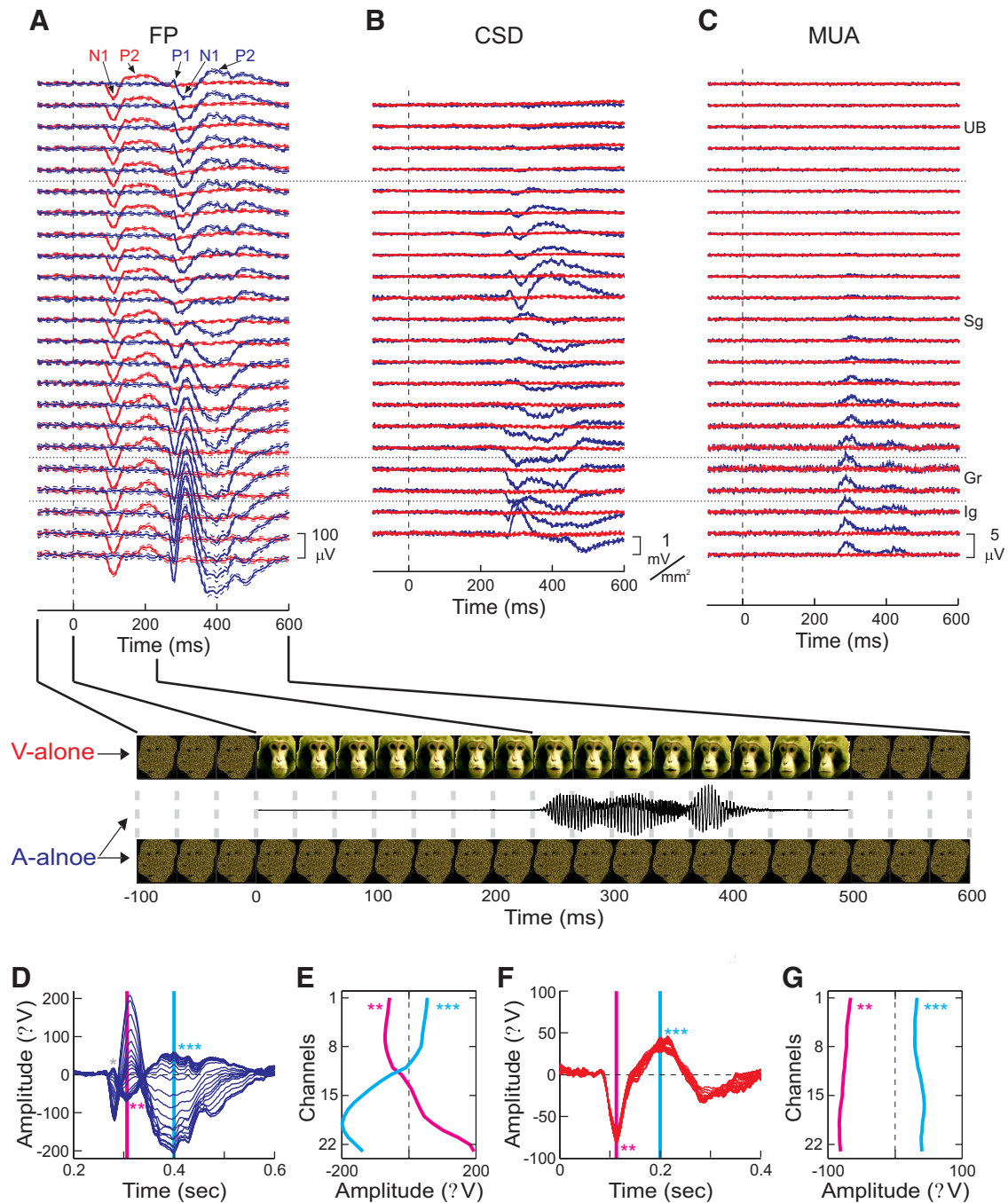


Figure 3. Representative FP responses to face and auditory stimuli in A1. **A**, Depth profile of FP responses to the face (red) and the sound (blue) of an exemplar AV vocalization during V-alone and A-alone trials of Task 1. Top to bottom, Positions of superficial to deep channels. Bottom inset, Expanded time courses of the visual and auditory stimuli. **B**, CSD responses. Downward (negative) and upward (positive) deflections are current sinks and sources, respectively. **C**, MUA responses. **A–C**, Lines indicate mean of 183 and 131 responses to visual and auditory stimuli, respectively. Dotted lines indicate the 95% CIs. Horizontal dashed lines indicate the borders between the upper bank of lateral sulcus (UB), supragranular layer (Sg), granular layer (Gr), and infragranular layer (Ig) in the auditory cortex. **D**, Superimposition of the auditory FP responses at all depths shown in **A**. Vertical bars represent the timing of two peaks in FP responses (magenta and cyan for N1 and P2 peaks, respectively). *P1, **N1, ***P2 in **D–G**. **E**, Amplitude distributions of auditory N1 and P2 across depths. **F**, Visual FP responses in same format as **D**. Between **D** and **F**, timing windows are different due to difference of the onset timing, 244 ms, between auditory and visual stimuli. However, durations of figures are same between **D** and **F**, to compare responses of two modalities at same resolutions. **G**, Amplitude distributions of visual N1 and P2.

Results

We trained 3 monkeys to perform tasks that require visual attention to face stimuli and auditory attention to vocal sounds (Fig. 1). Each trial entailed viewing a series of repeating video clips (duration 500 ms, between-movies interval 600–1200 ms) of vocalizing monkey face movies with (AV) or without vocalization sounds (visual or V-alone), or a static image with intermittent vocal

sounds (auditory or A-alone). Monkeys initiated a trial by pulling a lever, held the lever, and monitored repeatedly presented non-target stimuli until they detected a change in stimuli (i.e., a new face or voice in the sequence). During recording, the 3 monkeys (P, G, and W) responded to 0.47%, 0.40%, and 2.7% of nontargets, and 97.2%, 97.5%, and 91.4% of targets, respectively (median, $n = 37$, $n = 29$, and $n = 14$ sessions).

Different static images used for Tasks 1 and 2 created a difference in the behavioral context of the abrupt onset of faces (Fig. 1B). In Task 1, monkeys needed to monitor a repetition of the face onset to decide whether making a manual response or not. In Task 2, the face onset occurred only once in each trial and acted as an alarm to let monkeys prepare for upcoming stimuli. As we show below, responses to faces were observed when the abrupt onset of face occurred at the beginning of both movies during Task 1 and static images during Task 2.

Sensory responses in auditory cortex

In 3 monkeys (Monkeys P, G, and W) performing Tasks 1 or 2, recordings were made in the supratemporal plane (Monkey P: 45 sites; Monkey G: 33 sites; Monkey W: 17 sites); Figure 2A shows the positions of recording sites in Monkey P. Figure 3 shows responses to faces and vocal sounds at an A1 site during Task 1. Auditory responses started briefly after the onset of sound with the pattern of positive-negative-positive peaks at shallow depths (Fig. 3A, blue). We term those peaks as P1, N1, and P2, similarly to other studies (Donchin et al., 2001; Jellema et al., 2004; Riehle et al., 2013; Woodman et al., 2007; Bruyns-Haylett et al., 2017). Superimposition of FP responses concurrently recorded from all electrode channels outlines their amplitude and polarity changes with depth (Fig. 3D), as do the quantifications of peak amplitudes across depths (Fig. 3E). CSD derived from the FP responses depicted spatiotemporally distributed current sinks and sources (Fig. 3B, blue). MUA recorded concurrently with FP indicates robust local neuronal firing (Fig. 3C, blue). Most of MUA responses occurred below the depth of the inversion of FP responses (i.e., in the granular/infragranular layers). While MUA did not change in the top 12 channels, sink/source components were absent from only the top 4 channels of CSD profile, suggesting that the top 5 channels of the array electrode were above auditory cortex. The spatiotemporal patterns of FP and MUA responses and the presence of local CSD are typical of responses across classic sensory cortical regions (Schroeder et al., 1998; Lipton et al., 2010; Kajikawa and Schroeder, 2011; Kajikawa et al., 2015).

Similarly to previous studies (Hoffman et al., 2008), we also observed FP responses to faces in auditory cortex, with the patterns of negative-positive peaks, N1 and P2 (Fig. 3A, red). In contrast to auditory responses, FP responses to faces did not exhibit steep voltage gradients and polarity inversions across the depth of auditory cortex (Fig. 3F), were not accompanied by concomitant increase in MUA (Fig. 3C, red), and there were no underlying CSD responses (Fig. 3B, red). Similar spatiotemporal patterns of visual responses were observed at all other sites in the auditory cortex (see Fig. 6). While the lack of change in MUA alone still leaves a possibility of the presence of local subthreshold responses, the absence of CSD sinks/sources suggests that there were no local subthreshold events that could generate the observed FP responses either. A clue to the locations generating the face evoked FPs in auditory cortex is suggested by the components' laminar voltage gradients (Fig. 3G), in that FP responses to faces grew larger with depth at the site.

Across sites, median peak latencies of face-evoked N1 and P2 responses were 119.5 ms and 221.5 ms ($n = 58$) during Task 1. These values were similar to those of the N100 and P180 components reported previously (Hoffman et al., 2008). Hoffman et al. (2008) also showed that N1 component of the visually evoked LFP responses differed in peak amplitudes and latency between A1 and the middle lateral belt area (ML), and concluded that the visual responses were locally generated. In the present study, most of our recording sites during Task 1 were from areas A1 and

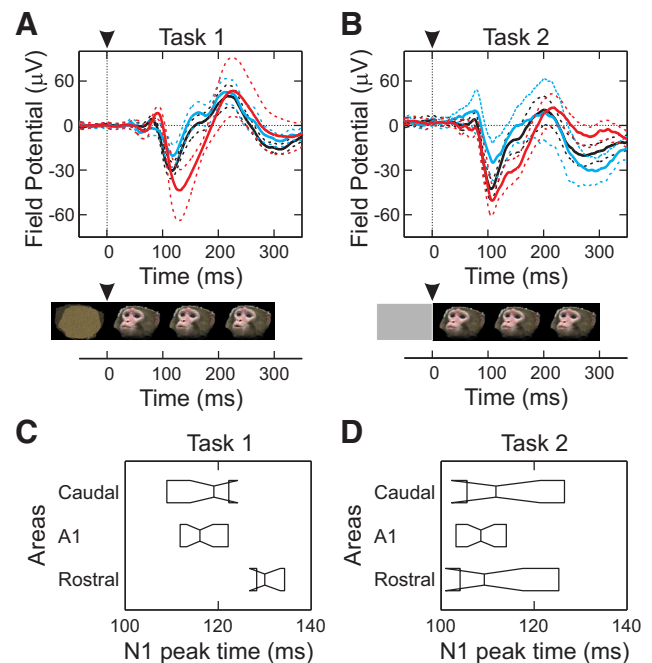


Figure 4. Comparisons of FP responses to faces between areas. **A**, FP responses to faces in V-alone trials during Task 1, averaged across sites in caudal areas (cyan, $n = 17$), A1 (black, $n = 33$), and rostral areas (red, $n = 8$). **B**, FP responses to the onset of static face images in all trials during Task 2, averaged across sites in caudal areas (cyan, $n = 12$), A1 (black, $n = 35$), and rostral areas (red, $n = 16$). **A**, **B**, Dotted lines indicate 95% CIs. Bottom insets, Stimulus time courses. Arrowheads indicate the timing of faces' appearance. **C**, Median and quartile of N1 peak latency of face responses in V-alone trials during Task 1 are plotted for three groups of areas. **D**, Median and quartile of N1 peak latency of visual responses to the onset of static face images during Task 2 are plotted for three groups of areas. **C**, **D**, Notches indicate 95% CIs of medians.

R in the core region, and from the caudal belt areas, and recordings in the lateral belt area were relatively sparse (Fig. 2). We sorted the recording sites into three groups along the caudal-rostral line: the caudal areas (CM, CL, and anterior Tpt), A1 (plus 2 ML sites), and the rostral areas (R and 2 rostral lateral belt area), and compared face-evoked FPs in the granular layers during Task 1 between the three groups (Fig. 4A). Median N1 peak latencies were 119 ms in A1, 116 ms in caudal areas, and 130 ms in rostral areas (Kruskal–Wallis, $\chi^2_{(2,55)} = 16.0$, $p = 3.3 \times 10^{-4}$; Fig. 4C). The N1 latency was significantly longer in rostral areas than in the other two groups (Tukey's HSD test, $p < 0.05$).

Longer peak latency could suggest different origins of visual N1 between the rostral areas and the other two groups. However, the spatiotemporal patterns of responses suggest otherwise. Figure 5 shows responses to faces and vocalizations at an R site during Task 1. Similar to the example shown in Figure 3, auditory FPs underwent clear polarity inversions across cortical layers, accompanied by robust CSD and MUA responses. FPs in area R also responded to faces with an N1 component peaking at 124.5 ms in the granular layer. Interestingly, the negative peak grew larger and peak timing shifted later gradually with depth. However, like the A1 case shown in Figure 3, FP responses to faces in R occurred without polarity inversions and concomitant CSD and MUA responses, even at the timing of late peak (Fig. 5B, dashed line), suggesting that difference in peak timing cannot be considered as reflection of local responses. It rather suggests that generators of visual FPs in R are activated later than those of visual FP in A1 or caudal areas.

Recording sites in different areal groups did not differ in the depth-dependent change in the FP amplitudes at the timing of N1 and P2. Thus, those groups are collapsed together to summarize

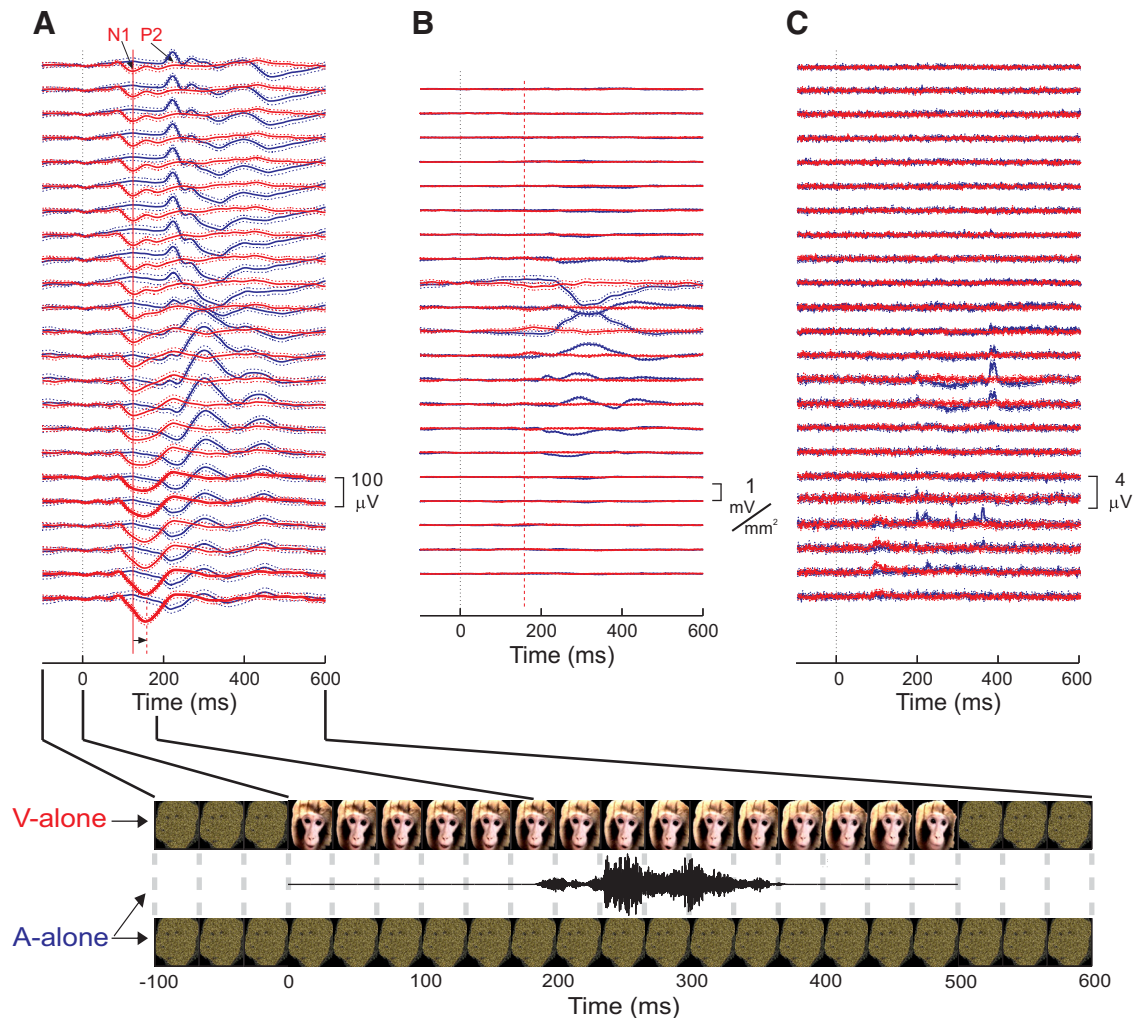


Figure 5. Representative FP responses to faces and vocalizations in area R. **A**, Depth profile of FP responses to the face (red) and the sound (blue) of an exemplar AV vocalization during V-alone and A-alone trials of Task 1. Intercontact spacing on the electrodes was $200\ \mu\text{m}$. Red vertical line indicates the visual N1 peak timing, 124.5 ms, at a channel in layer 4. Short red vertical dashed line with a black horizontal arrow indicates that N1 peak timing shifted to 158.5 ms in the deepest channel located in white matter. Bottom inset, Expanded time courses of the visual and auditory stimuli. **B**, CSD responses. Downward and upward deflections are current sinks and sources, respectively. Red vertical dashed line indicates 158.5 ms. **C**, MUA responses. **A–C**, Lines indicate mean of 146 and 78 responses to visual and auditory stimuli, respectively. Dotted lines indicate the 95% CIs.

the general patterns of the amplitude distributions of sensory responses against depth (Fig. 6). Across sites, auditory FP responses consistently exhibited consistent polarity inversions over depth in auditory cortex, accompanied by phasic elevations of MUA (Fig. 6A). Face responses had distinctly different laminar distributions (Fig. 6B); the N1 and P2 components gradually increased their peak amplitudes with depth but showed no abrupt changes in polarity, and we observed no concomitant MUA responses.

Sensory responses to nontarget stimuli described so far were recorded when monkeys performed Task 1, in which animals had to keep track of stimuli to detect oddballs and decide whether to make a manual response or not after each stimulus. However, in previous studies of visual responses in auditory cortex, animals simply had to maintain fixation without any need to attend or to make sensory discriminations. Thus, behavioral requirements of Task 1 differed from those in previous studies (Ghazanfar et al., 2005; Hoffman et al., 2008). Monkeys in the present study were also trained to perform Task 2, in which the onset of face on the screen occurred only once in the beginning of each trial (Fig. 1). At that time, the behavioral condition requiring animals to only maintain gaze with no need of sensory tracking or manual responses was similar to the condition used in previous studies.

Figure 4B shows layer 4 FP responses to the onset of static face images during Task 2 in caudal, A1, and rostral areas. Even under the condition requiring only maintenance of gaze, FP responses to faces occurred across all three grouped areas. Median N1 peak latencies were similar across groups: 108.5 ms in A1, 112 ms in caudal areas, and 109 ms in rostral areas (Kruskal–Wallis, $\chi^2_{(2,60)} = 1.24$, $p = 0.538$; Fig. 4D). Median N1 and P2 latencies, collapsing the three groups together, were 107 and 195 ms. Spatial patterns of N1 and P2 component amplitudes in response to nontarget vocal sounds and static face images during Task 2, collapsing data across the three areal groups, are summarized in Figure 6C, D. Like the corresponding components recorded during Task 1, visual FP components in auditory cortex during Task 2 did not exhibit polarity inversions but grew in amplitude with increasing depth.

Sensory responses below auditory cortex

The spatial gradient of the visual FPs suggests that their generators are located below the auditory cortex. We tracked sensory responses below auditory cortex while animals repeatedly performed blocks of same tasks. Between blocks, electrode arrays ($200\ \mu\text{m}$ interchannel spacing) were shifted in 3–4 mm steps to systematically map responses over depth. Figure 7A shows FPs

recorded during a 1.5 s period that included the onset of face image and the first nontarget A-alone stimulus in Task 2, at 3 successive recording array depths. Total sampled distance was ~ 12 mm.

Recordings during the first block (approximately top one-third) straddled auditory cortex, as seen in the inversion of auditory FP responses (Fig. 7A, arrowhead). In the array positions below that inversion, auditory FP responses gradually became smaller and underwent another rapid shift in spatial gradient (Fig. 7C, double arrowhead) corresponding to the depth of the supragranular layers in the superior temporal polysensory area (STPa), in the upper bank of the superior temporal sulcus (STS); this effect occurred with concomitant MUA in 7 of 37 the penetrations in 3 monkeys (see Fig. 11), although it is not well captured in the quantification of the amplitude of FP (Fig. 7D). This is because the inversion of relatively weak auditory FP responses in STPa is superimposed on larger FP responses that are volume-conducted from the overlying auditory cortex. However, superimposed curves of CSD amplitude indicate the presence of local auditory responses in STPa.

The face onset-evoked N1 is discernible at all depths (Fig. 7B). As depth increases, the amplitudes of N1 and P2 become larger. Although the polarity of component peaks does not change appreciably down to the depth of the STS, there is an abrupt shift in the amplitude gradient and a polarity inversion at 7.6 mm below the inversion of auditory FP responses (Fig. 7D). In addition, a small positive peak (P1) before N1 prevailed above the visual FP inversion, and this also inverted to become a negative peak at lower depths (Fig. 7B, bottom, asterisk, labeling the negative counterpart).

Similar depth patterns of visual and auditory FP responses were observed in all experiments in the 3 monkeys ($n = 13$ in Monkey P and $n = 14$ in Monkey G during Task 2). The median distance between the inversions in the auditory and the visual FP profiles was 8.0 mm (range: 6.6–11.4 mm). Because the N1 inversions occur in the supragranular layers of the sensory cortices, these values approximate the distances between the supragranular layers of areas in the lower bank of the lateral sulcus and those in the lower bank of the STS. Even though there could be an overestimate of the distance in cases with mildly oblique penetration angles, these values are consistent with the range of distance between the lateral sulcus and the STS measured on the brain surface over the range of the AP range examined in the present study (Kajikawa et al., 2015), and those from a standard macaque brain atlas (Paxinos et al., 2008; Saleem and Logothetis, 2012). In addition, smooth spatial gradient of peak amplitude with stable peak timing over distances across gray matter, white matter, and pia in the STS suggest that the effect of spatially nonuniform tissue conductivity and permittivity is minor (see also Kajikawa

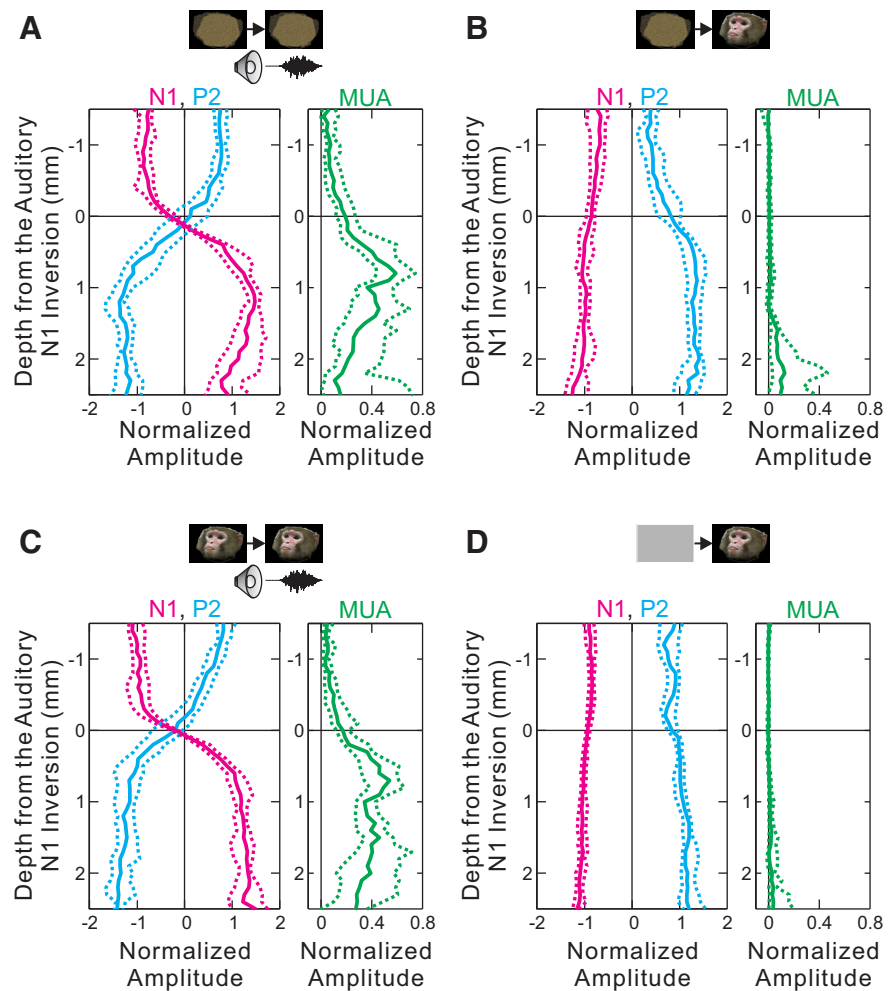


Figure 6. Depth-amplitude/polarity gradients for N1 (magenta), P2 (cyan), and MUA (green) in auditory cortex. **A**, Normalized depth-amplitude/polarity distributions of auditory N1, P2, and MUA responses to sounds during A-alone trials in 3 monkeys performing Task 1 ($n = 16$, $n = 27$, and $n = 15$); median amplitude values with 95% CIs (bootstrap, 1000 resampling) are plotted against depth relative to that of the inversion of auditory N1. **B**, Normalized depth-amplitude/polarity distributions of N1, P2, and MUA of visual face-onset responses during V-alone trials at the same sites as those in **A** during Task 1. **C**, Normalized depth-amplitude/polarity distributions of auditory N1, P2, and MUA responses during A-alone trials in 3 monkeys performing Task 2 ($n = 29$, $n = 29$, and $n = 5$). **D**, Normalized depth-amplitudes/polarity distributions of N1, P2, and MUA of visual responses to the onset of static face images in all trials at the same sites as those in **C** during Task 2.

and Schroeder, 2011, 2015). Figure 7F summarizes the depth patterns of the peak amplitudes of visual and auditory N1, revealing that inversions occurred at different depths consistently across penetration sites. N1 of visual FP responses was significant at all depths above IT. Figure 7E shows similar responses across depth during Task 1 ($n = 15$ in Monkey G and $n = 2$ in Monkey W). The depths below STPa indicated that the inversions of FP responses to faces occurred in the lower bank of STS, corresponding to the IT cortex; penetration tracks reaching IT were confirmed histologically (Fig. 2B).

Spatiotemporal profiles of visual responses in IT

The spatiotemporal profile of FP responses to faces in IT is shown in Figure 8A. In the superficial recording sites, located in the supragranular layers of IT and in the overlying cortical area (TPO), the FP components have the same positive-negative-positive pattern (P1, N1, P2, labeled by vertical gray, magenta, and cyan lines, respectively, in Fig. 8C) that extends upward, with gradual amplitude decay into auditory cortex, and above (Fig. 7). All these components undergo polarity inversion across the lay-

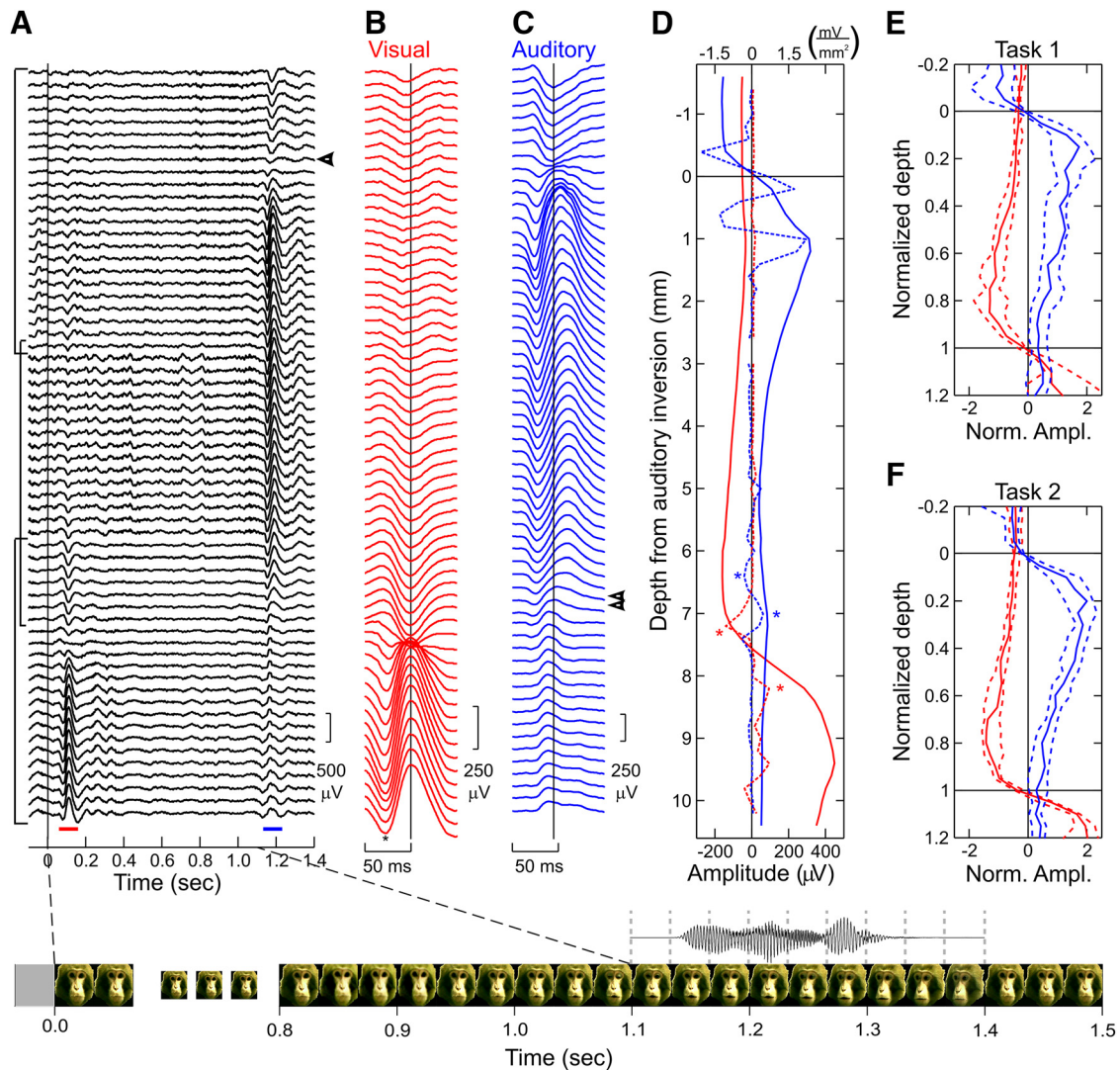


Figure 7. Tracking FP responses below auditory cortex. **A**, A depth FP profile at 0.2 mm resolution from auditory cortex (top) to a point ~ 10 mm below auditory cortex (bottom) during A-alone trials in Monkey P performing 3 blocks of Task 2. The depth of the electrode array was shifted between blocks: 3 brackets at the left margin circumscribe electrode channels recorded simultaneously within each block. At depths where electrode array positions were overlapping, FPs corresponding to the overlap positions were averaged. Arrowhead indicates the depth of the polarity inversion of the auditory FP responses. Time course of stimuli at the bottom indicates that a still face image appeared at 0 ms, followed by a movie clip starting at 900 ms, and then by a vocal sound with an additional 244 ms delay. **B**, Expanded view of the FP during the 100 ms period indicated by red horizontal bar in **A** (bottom). Vertical line indicates the timing of the visual N1 component. *Visual P1. **C**, Expanded view of the FP during the 100 ms period indicated by blue horizontal bar in **A** (bottom). Vertical line indicates the timing of the auditory N1. Double arrowheads indicate the depth of STPa where auditory FP responses show an additional abrupt shift in their laminar voltage gradient, albeit weaker than the frank polarity inversion seen in auditory cortex. **D**, Amplitude/polarity of visual (red) and auditory (blue) N1 components in **B** and **C** plotted against the distance from the auditory N1 inversion. Superimposed dashed lines indicate the amplitudes of visual (red) and auditory (blue) CSD at the timing of N1. *Supragranular sink and source pairs in IT (red) and STPa (blue). Slight overlap of CSD components between STPa and IT is due to coarse spatial sampling with 200 μm intercontact intervals. **E**, Median and 95% CIs ($n = 17$) of N1 amplitudes during Task 1 across depth. Amplitudes were normalized within each penetration track, separately for auditory and visual N1. Depth is calibrated relative to the distance between depths of auditory and visual N1 inversions. **F**, Same as **E** for FP responses during Task 2 ($n = 27$).

ers of IT (Fig. 8C,D). Figure 8B shows the CSD profile derived from the FP profile and selected channels from the MUA profile concurrently recorded with FP. The largest CSD components and the large amplitude MUA are found at and below the depth of the FP inversion (Fig. 8B). These spatial patterns were consistent with the well-known laminar patterns of FP, CSD, and MUA responses to sensory stimuli in other sensory cortices: starting from feedforward activation of the granular layer followed by activation of extragranular layers (Schroeder et al., 1991, 1995; Kajikawa and Schroeder, 2011; Lipton et al., 2010).

Figure 9 summarizes depth-amplitude profiles of visual and auditory FP responses in IT, plotting the median values for normalized amplitudes of P1, N1, and P2, and MUA response magnitude against the depth relative to the inversion of visual N1. All

of these FP components underwent polarity inversion across the layers of IT during Task 2 (Fig. 9B). The P1 inversion (Fig. 9B, small arrow) was significantly deeper than the N1 and P2 inversions, outlining a local P1 generator in the granular layer along with the main generators of both N1 and P2 in the supragranular layers. The most prominent MUA responses occurred below the inversions of visual FP components in the granular and infragranular layers. In contrast, none of the auditory FP components showed a local polarity change in IT, and none had a local MUA correlate (Fig. 9A). During Task 1, although CIs were large due to small number of penetration sites, features of face and auditory responses (Fig. 9C,D) were similar to those during Task 2.

If FP responses to faces in auditory cortex were volume-conducted from IT, then FP responses in the vicinity of IT should

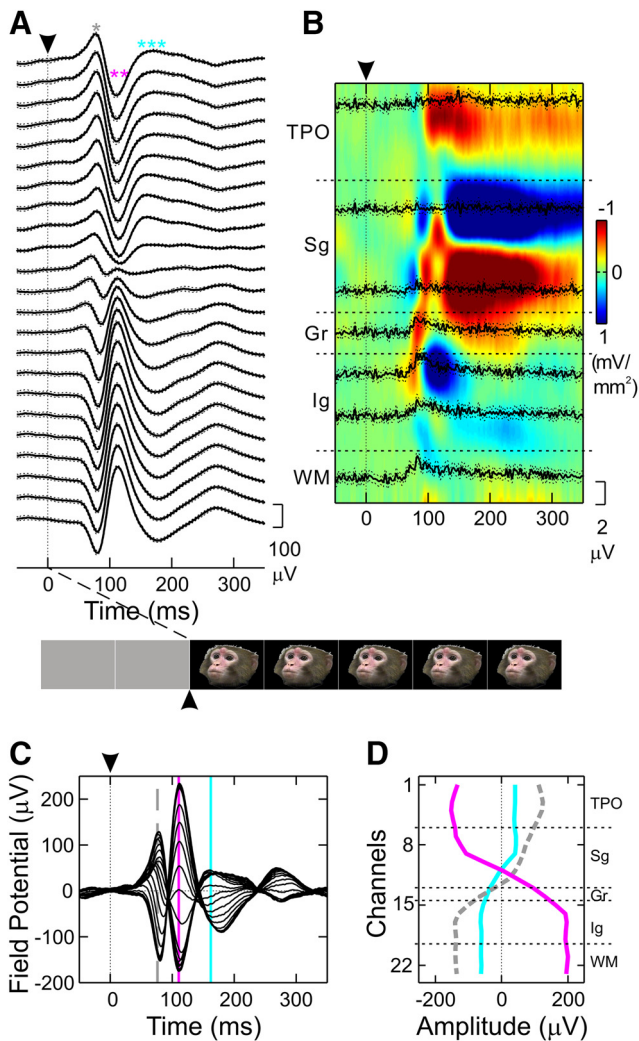


Figure 8. A representative laminar profile of face responses in IT. **A**, Mean FP responses to static face images at the beginning of all trials ($n = 86$), simultaneously recorded across the layers of IT, in Monkey G performing Task 2 using an electrode with $200 \mu\text{m}$ intercontact spacing. The electrode was positioned to bracket the layers of IT, so the uppermost channels were located in the supragranular layers of the overlying area TPO and the lowermost channels were in the white matter below IT. Bottom inset, Stimulus time course. **A–C**, Arrowheads indicate the timing of screen transition. *P1. **N1. ***P2. **B**, Mean CSD and MUA responses (color plot and line traces, respectively) to faces simultaneously recorded with FP shown in **A**. **A**, **B**, Dotted lines indicate 95% CIs. **C**, Superimposition of the FP responses in **A**. Vertical lines indicate the timing of P1 (gray dashed), N1 (magenta), and P2 (cyan). **D**, Quantified depth distributions of P1, N1, and P2 peak amplitudes.

have the peaks at timing similar to those in auditory cortex. Indeed, median peak latencies of N1 and P2 of visual responses: 126 ms and 192.5 ms ($n = 20$) during Task 1 and 117.5 ms and 167.5 ms ($n = 35$) during Task 2 were indeed similar to those in auditory cortex (see above). In combination with the clear pattern of gradually increasing component amplitudes tracking from auditory cortex down to IT, these data suggest that visual FPs in IT contribute to those recorded in auditory cortex. This inference is further explored by more quantitative methods below.

Local intracortical contributions to sensory FP responses

To test whether the local CSD could generate FP responses, we used volume conductor modeling to calculate the “predicted” spatiotemporal profiles of FP responses (i.e., those that could be generated by the empirically derived CSD profiles) and to exam-

ine their similarity to the directly “observed” FP profiles (Kajikawa and Schroeder, 2015). Figure 10A, B shows visual FP and CSD responses to faces in a representative IT site. The CSD response began with a sink at the granular/infragranular boundary, followed by another sink in supragranular layers (Fig. 10A). Spatial integration of the CSD by the model produced a spatiotemporal pattern of FP similarly to that of the recorded FP, and two profiles contained common peaks (compare Fig. 10A, C).

In auditory cortex, CSD of visual responses yields very low amplitude sink-source configurations with no apparent correspondence to the laminar FP profile (Fig. 10E); this likely reflects the average of ongoing activity unrelated to the visual response seen in the FP. Accordingly, integration of these CSD components results in spatially localized negative and positive components of FP (Fig. 10F), which appear unrelated to the observed visual FP responses (Fig. 10D). Generally, among sites in auditory cortex, application of the volume conductor model to visual responses creates predicted profiles that are weak and disorganized, with little or no correspondence to the observed FP profile. Thus, local auditory cortical activity does not appear to account for the locally recorded visual FPs (i.e., these FPs are volume-conducted from nonlocal sites). In the same AC sites, auditory CSD responses (Fig. 10H) can generate auditory FP responses with a spatiotemporal pattern comparable with that of the observed FP responses (Fig. 10G, I).

The similarity score defines the correspondence of the spatiotemporal profile of the model-derived (predicted) FP responses to that of the recorded (observed) FP responses (Kajikawa and Schroeder 2015). These were compared across visual responses in IT and auditory cortex, and auditory responses in auditory cortex. Median similarity was as follows: 0.82, 0.10, and 0.72 (Kruskal–Wallis, $\chi^2_{(2,95)} = 66.8$, $p = 3.1 \times 10^{-15}$) for responses during Task 1 (Fig. 10J), and 0.78, 0.21, and 0.81 ($\chi^2_{(2,154)} = 82.75$, $p = 1.1 \times 10^{-18}$) during Task 2 (Fig. 10K). In both tasks, the similarity scores of face responses in auditory cortex are significantly lower than those of face responses in IT and auditory responses in AC. These results provide no evidence that the generators of FP responses to face in auditory cortex were generated locally. Rather, it appears that they are volume-conducted from one or more remote locations, such as IT.

STPa

STPa lies between IT and AC, occupying the upper bank of STS, and is responsive to both visual and auditory inputs (Bruce et al., 1986; Schroeder and Foxe, 2002; Dahl et al., 2009), some of which may be selective to faces (Baylis et al., 1987). In the present study, we rarely observed bimodal STPa sites but commonly found sites dominated by one or the other modality, in keeping with the anatomical connections patterns of the region (Seltzer et al., 1996; Cusick, 1997). Monkeys G and W had visually evoked MUA responses in STPa in 13 of 20 and 1 of 3 sites, and none of them responded to auditory stimuli. STPa of the Monkey P had auditory MUA responses in 7 of 14 sites. Two of 7 sites also had suppressive visual responses to face stimuli.

Generation of visual ERP may receive contributions from visually responsive STPa sites. Figure 11A shows responses along a track in which both IT and STPa had excitatory responses. The CSD profile in STPa has a spatial pattern basically upside-down relative to that in IT; this is predictable because these areas are opposed along their pial surfaces. Therefore, STPa activity may generate patterns of FPs opposite in polarity to those of IT. However, most of the clear inversions of visual FP responses in the vicinity of STS occurred in IT (Fig. 11A, left). The laminar profile

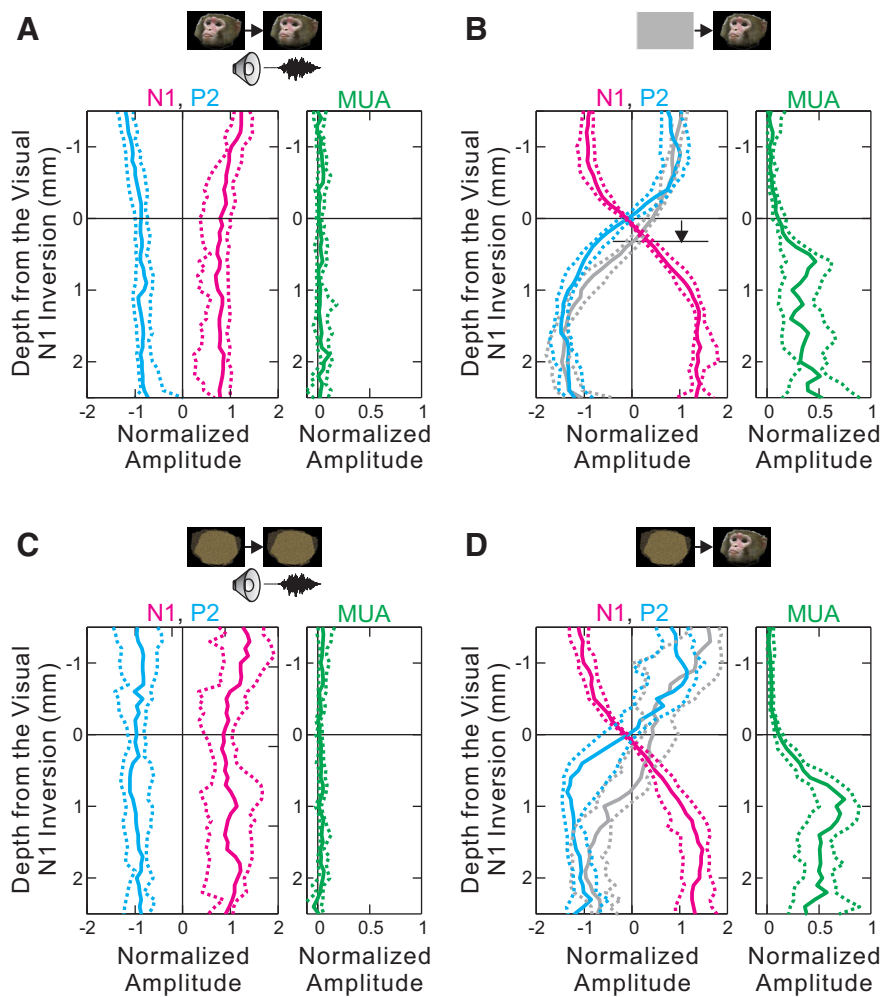


Figure 9. Depth distributions of normalized amplitudes of auditory (**A, C**) and face responses (**B, D**) in IT. **A**, Normalized depth-amplitude/polarity distributions of auditory N1 (blue), P2 (magenta), and MUA (green) responses to sounds of A-alone trials in IT of 3 monkeys performing Task 2 ($n = 14$, $n = 20$, and $n = 1$); median amplitude values with 95% CIs (bootstrap, 1000 resampling) are plotted against depth relative to that of the inversion of visual N1 in **B**. **B**, Normalized depth-amplitude/polarity distributions of P1 (gray), N1 (magenta), P2 (blue), and MUA (green) for face responses to static face images at the beginning of all trials during Task 2 at the same sites as those in **A**. **C**, Normalized depth-amplitude/polarity distributions of auditory N1, P2, and MUA responses to sounds of A-alone trials in IT of 2 monkeys performing Task 1 ($n = 17$ and $n = 3$). **D**, Normalized depth-amplitude/polarity distributions of P1, N1, P2, and MUA of face responses in V-alone trials during Task 1 at same sites as those in **C**.

of polarity inversion in IT was consistent across all tracks that had excitatory responses in both IT and STPa. Figure 11*B* shows responses along a track in which STPa has no visual responses, but strong excitatory auditory responses (Fig. 11*C*). Together, in all tracks, the N1 component of visual responses generated in IT remains visible beyond the upper bank of STS regardless of STPa activity.

Postmortem histology, available from Monkey P, showed that all tracks with the inversion of visual FP responses reached the lower bank of STS (Fig. 2*B*). Overall, these results support the conclusion that visual FP responses to faces in the temporal region of macaques is mostly generated in IT cortex, and is volume-conducted through intervening brain regions, including the auditory cortices, up to the brain surface and, ultimately, to the scalp.

Discussion

We observed FP responses to faces in auditory cortex, as reported previously. Although they are reliable, they do not exhibit a local

laminar voltage gradient (e.g., polarity inversion) suggestive of a local contribution to the FP. Consequently, CSD responses to faces in auditory cortex were almost flat, providing no evidence that auditory cortex contributes to the observed spatio-temporal patterns of FP responses to faces. This pattern of results is in stark contrast to local auditory responses, which have laminar FP profiles that exhibit clear polarity inversions across depth and have underlying CSD configurations that clearly can generate the observed FP responses. The gradual increase in the peak amplitudes of FP responses to faces over depth in auditory cortex suggests that these FPs are generated below auditory cortex. As we tracked down the face responses below auditory cortex, we found clear and consistent polarity inversion of visual FP responses in IT. Analysis of the local FP generator patterns with volume conductor modeling provided quantitative confirmation that the polarity-inverting FP components were locally generated in the granular and supra-granular layers of IT. Although there is a suggestion that STPa, an area lying between IT and auditory cortex in the upper bank of the STS, may also contribute to the generation of visual FPs that appear in the auditory cortex, local generation of visual FPs in STPa is weaker and inconsistent across sites. The overall pattern of results argues strongly that FP responses to faces in auditory cortex were most likely generated in IT and volume-conducted to auditory cortex.

Local field or far field

The term LFP assumes that the events generating LFP occur in the vicinity of the recording site. However, this is not a given because LFPs clearly can be generated by remote loci and volume-conducted to the recording site. The benefit of volume conduction is that FPs generated by synaptic currents within the brain can be recorded non-invasively at the scalp in humans. The drawback is that, even when recording within the brain, signals can be either contaminated by far FPs or solely composed of far FPs. This is of particular concern for studies that propose to assign specific components of FPs, such as α or γ oscillations, to specific laminar divisions (e.g., the supra-granular or infragranular layers) of the neocortex (Haegens et al., 2015).

When FPs are correlated with neuronal firing, they are more likely to reflect local activity. In the absence of local neuronal firing, FPs could still reflect subthreshold synaptic currents in local neurons. In either case, simultaneous multisite recordings distinguish local and far-field activity. In such a recording, far-field FPs have gradual changes in amplitude across adjacent sites without changes in neuronal firing. On the contrary, genuinely local responses would have a steep gradient and polarity inversion, often with concomitant neuronal firing. It may be noted

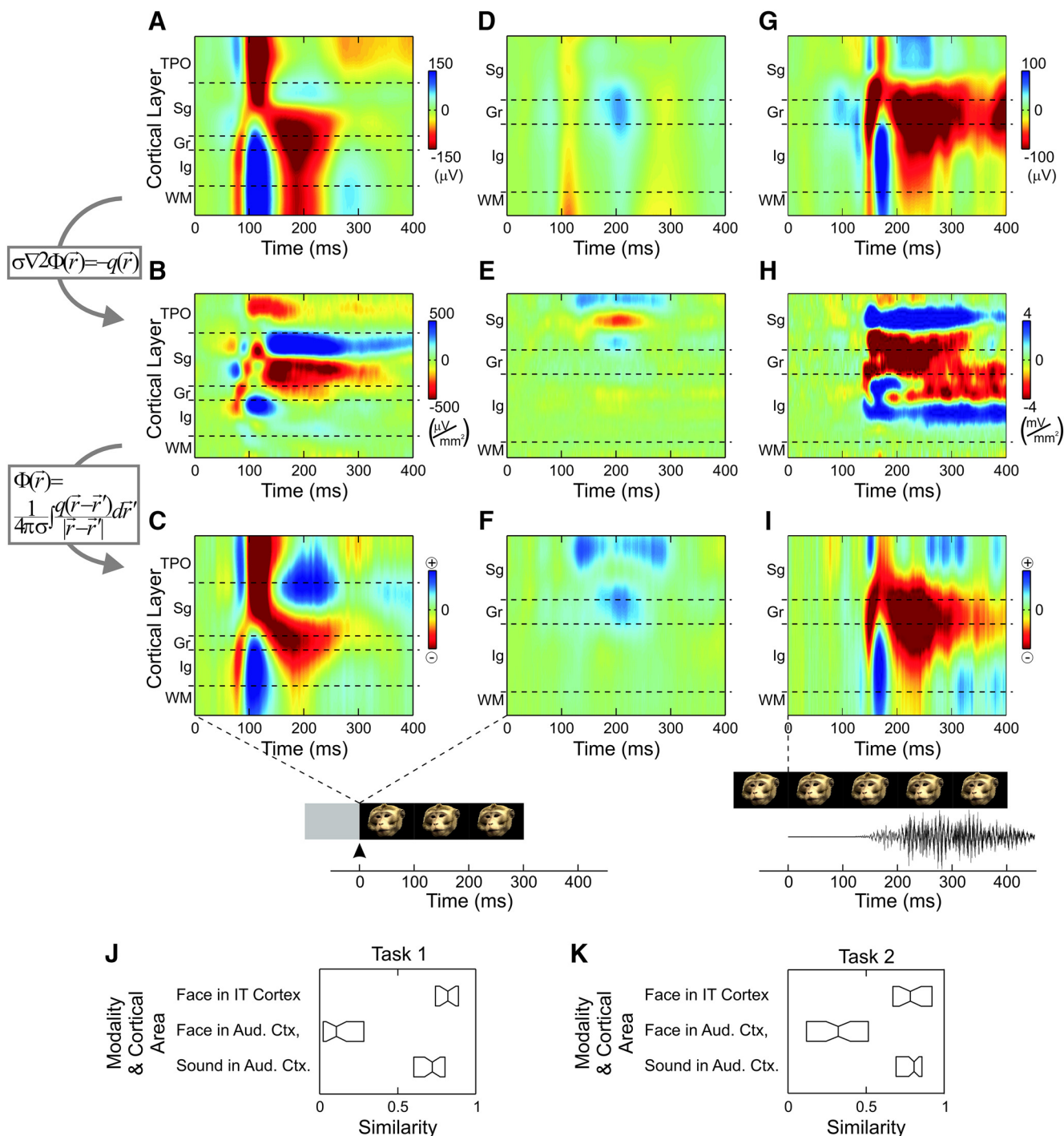


Figure 10. Quantifying the relative contribution of local activity to FP responses in IT and AC. **A**, Spatiotemporal (color map) profile of mean FP response to static face image at the beginning of all trials ($n = 86$) during Task 2 in an IT site. **B**, CSD analysis of the FP profile in **A**. **C**, FP derived from CSD in **B** by volume conductor modeling. **D–F**, Spatiotemporal profiles of mean FPs and CSD responses to stimuli of V-alone trials ($n = 205$) during Task 2 in an auditory cortical site, in same formats as **A–C**. **G–I**, Spatiotemporal profiles of FP and CSD responses to a vocal sound of A-alone trials ($n = 16$) during Task 2 in the same site as **D–F**. **D–F**, **G–I**, The rows have common color scales. Bottom insets, Time courses of stimuli. **G–I**, The onset of sound was 137.6 ms. **J**, Median and quartile of similarity scores between the spatiotemporal patterns of the recorded (observed) and the model-derived (predicted) FP responses during Task 1 are plotted for face response in IT ($n = 20$) and face responses and auditory responses in auditory cortex ($n = 39$). **K**, Same as **J** for the face response in IT ($n = 31$) and face responses and auditory responses in auditory cortex ($n = 63$) during Task 2.

that putting a reference electrode close enough to the recording site would automatically eliminate most of far-field components (Kajikawa and Schroeder, 2011). Although this quantity better approximates a “local” FP, it might be more appropriate to call it the “potential gradient” or “field strength” rather than “potential.”

Given spatially uniform FP-like face responses of FP in auditory cortex, one may still force the idea of local origins and derive the underlying CSD using the inverse CSD (iCSD) method (Pettersen et al., 2006; Einevoll et al., 2013). Indeed, Hunt et al. (2011) applied such local origin modeling of iCSD to uniform FPs recorded subcortically in rats and obtained CSD consisting of a

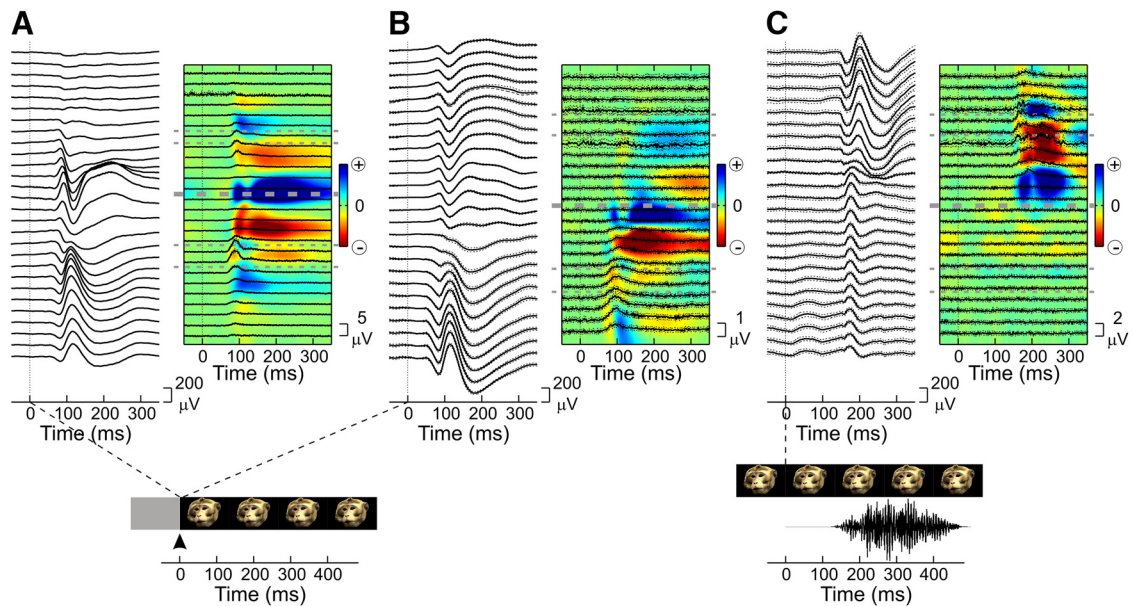


Figure 11. Sensory response profiles in STPa. **A**, Spatiotemporal profiles of mean FP (left) and CSD (right) responses to static image at the beginning of all trials during Task 2 at 0.2 mm intervals from STPa (higher) to IT (lower) recorded from a penetration in Monkey G. Concomitant MUA responses at corresponding depths are superimposed on the CSD plots. STPa and IT both responded with robust excitation. Bottom insets, Stimulus time courses. Inset, Arrowhead indicates the time when a static face image appeared in the rectangle window. **B**, FP and CSD responses to static image at the beginning of all trials during Task 2 along a track bracketing STPa and IT in Monkey P. **C**, Auditory-evoked FP and CSD responses to a vocal sound of A-alone trials during Task 2 along the same track as **B** in Monkey P. The onset of sound was at 137.6 ms. Robust auditory responses were confined to STPa. **A–C**, Dotted lines for both FP and MUA indicate 95% CIs.

spatially uniform monopole. A spatially uniform monopole could certainly generate spatially uniform FP-like face responses we have shown for the macaque auditory cortex. However, the presence of such neuronal processes that elongate several millimeters and cross cortical borders perpendicularly is highly unlikely. Thus, iCSD could result in model-dependent artificial CSD. The optimal use of iCSD in such a case could be to model a remote source of CSD at the loci where neuronal firing occurred. However, such models still require determination of the distance to the source from the recording sites, and cannot estimate the CSD components generating the FP of opposite polarity that are located beyond the site of the polarity inversion. Thus, spatial tracking of FP would be a better approach to identify the genuine generators of far FPs.

Generator of face responses in the macaque temporal lobe

Previous studies of LFP responses to face in auditory cortex did not address spatial patterns of LFPs but rather assumed responses were local (Ghazanfar et al., 2005; Hoffman et al., 2008; Kayser et al., 2008). However, we find that the spatial topography of FP responses to faces and sounds in auditory cortex appear exactly as predicted for far and local fields, respectively. The components (N1 and P2) of visual FP responses do not change their timing appreciably and continue to appear at sites that are in the white matter below auditory cortex. Similar peak timing and constant polarity were expected features of volume conduction (Nunez and Srinivasan, 2006; Cohen et al., 2009; Kajikawa and Schroeder, 2011). Thus, our results raise the possibility that that previously described face LFP responses to face in auditory cortex are also far-field reflections of FPs generated in distant sites.

Our data also show that signatures of local events generating visual responses do appear, albeit inconsistently, in areas just above the STS, such as STPa. Other studies have shown that LFP responses to faces or other visual stimuli in structures above the STS occurred consistently with negative peaks at 100–120 ms

similar to N1 (Anderson et al., 2008; Matsuo et al., 2011; Tureson et al., 2012), and that positive peaks of similar timing occurred within IT (Woloszyn and Sheinberg, 2009; Meyer and Olson, 2011). Thus, those results and ours concur in the suggestion that much of the LFP response to faces in the macaque temporal lobe above STS, with the exception of parts of STPa, are not locally generated but rather are volume-conducted from below the STS.

Finding local generators of visual responses in the lower bank of STS is consistent with the extant literature on face processing (Gross et al., 1972; Perrett et al., 1982; Desimone et al., 1984; Baylis et al., 1987; Tanaka et al., 1991; Sugase et al., 1999; Pinsk et al., 2005). IT contains several patches that are face selective (Tsao et al., 2008; Freiwald and Tsao, 2010). Because we did not examine responses to a wide variety of visual stimuli, such as faces, other body parts, places, etc., we could not define the face selectivity of our recording sites. Even though behavioral conditions in studies of IT, visual fixation, and passive viewing are similar to Task 2, we observed slower visual responses in rostral areas than A1 or caudal areas during Task 1. The results are consistent with gradual increase in the latency of face responses from posterior to anterior face responsive regions of IT (Freiwald and Tsao, 2010).

One feature of visual responses in the lower bank of STS that has not been shown before was the spatiotemporal profiles of CSD responses. As we show, the CSD profile of visual responses in IT started from the granular layer and propagated to extragranular layers. The pattern was quite similar to those shown in other cortical areas (Schroeder et al., 1995; Kajikawa and Schroeder, 2011), and consistent with laminar structure of inputs to IT (i.e., it receives afferents into the middle layers from lower visual cortex) (Seltzer and Pandya, 1989; Distler et al., 1993; Saleem et al., 1993; Ungerleider et al., 2008).

Our findings show that far-field FPs can be almost as strong as locally generated FPs. This may disprove the prior conclusion that FP responses to face arise locally in the macaque auditory

cortex (e.g., Hoffman et al., 2008). Given that prior studies visual FPs in auditory cortex motivated this one, we did not study responses to a battery of visual stimuli. However, the effects of volume conduction on visual FPs recorded in auditory cortex are unlikely to be specific to face-evoked responses (i.e., any visual stimuli that strongly activates IT may evoke far-field visual FP responses in the auditory cortex). Regardless, our data do not dispute the observation of nonauditory responses in auditory cortex, when they manifest in neuronal firing (Brosch et al., 2005; Bizley et al., 2007) or CSD (Lakatos et al., 2009) because those signals are reliable indicators of local activity. However, when FP responses are observed without evidence of simultaneous firing or CSD (Kral et al., 2003), they are subject to the concern that their generators lie at some distance from the recording site.

References

- Anderson B, Mruczek RE, Kawasaki K, Sheinberg D (2008) Effect of familiarity on neural activity in monkey inferior temporal cortex. *Cereb Cortex* 18:2540–2552. [CrossRef Medline](#)
- Baylis GC, Rolls ET, Leonard CM (1987) Functional subdivisions of the temporal lobe neocortex. *J Neurosci* 7:330–342. [Medline](#)
- Bizley JK, Nodal FR, Bajo VM, Nelken I, King AJ (2007) Physiological and anatomical evidence for multisensory interactions in auditory cortex. *Cereb Cortex* 17:2172–2189. [CrossRef Medline](#)
- Brosch M, Selezneva E, Scheich H (2005) Nonauditory events of a behavioral procedure activate auditory cortex of highly trained monkeys. *J Neurosci* 25:6797–6806. [CrossRef Medline](#)
- Bruce CJ, Desimone R, Gross CG (1986) Both striate cortex and superior colliculus contribute to visual properties of neurons in superior temporal polysensory area of macaque monkey. *J Neurophysiol* 55:1057–1075. [Medline](#)
- Bruyns-Haylett M, Luo J, Kennerley AJ, Harris S, Boonman L, Milne E, Vautrelle N, Hayashi Y, Whalley BJ, Jones M, Berwick J, Riera J, Zheng Y (2017) The neurogenesis of P1 and N1: a concurrent EEG/LFP study. *Neuroimage* 146:575–588. [CrossRef Medline](#)
- Buzsáki G, Anastassiou CA, Koch C (2012) The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. *Nat Rev Neurosci* 13:407–420. [CrossRef Medline](#)
- Chandrasekaran C, Lemus L, Trubanova A, Gondan M, Ghazanfar AA (2011) Monkeys and humans share a common computation for face/voice integration. *PLoS Comput Biol* 7:e1002165. [CrossRef Medline](#)
- Chandrasekaran C, Lemus L, Ghazanfar AA (2013) Dynamic faces speed up the onset of auditory cortical spiking responses during vocal detection. *Proc Natl Acad Sci U S A* 110:E4668–E4677. [CrossRef Medline](#)
- Cohen JY, Heitz RP, Schall JD, Woodman GF (2009) On the origin of event-related potentials indexing covert attentional selection during visual search. *J Neurophysiol* 102:2375–2386. [CrossRef Medline](#)
- Cusick CG (1997) The superior temporal polysensory region in monkeys. In: *Extrastriate cortex in primates* (Rockland KS, Kaas JH, Peters A, ed), pp 435–468 New York: Plenum.
- Dahl CD, Logothetis NK, Kayser C (2009) Spatial organization of multisensory responses in temporal association cortex. *J Neurosci* 29:11024–11032. [CrossRef Medline](#)
- Desimone R, Albright TD, Gross CG, Bruce C (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci* 4:2051–2062. [Medline](#)
- Distler C, Boussaoud D, Desimone R, Ungerleider LG (1993) Cortical connections of inferior temporal area TEO in macaque monkeys. *J Comp Neurol* 334:125–150. [CrossRef Medline](#)
- Donchin O, Gribova A, Steinberg O, Bergman H, Cardoso de Oliveira S, Vaadia E (2001) Local field potentials related to bimanual movements in the primary and supplementary motor cortices. *Exp Brain Res* 140:46–55. [CrossRef Medline](#)
- Driver J, Noesselt T (2008) Multisensory interplay reveals crossmodal influences on ‘sensory-specific’ brain regions, neural responses, and judgments. *Neuron* 57:11–23. [CrossRef Medline](#)
- Einevoll GT, Kayser C, Logothetis NK, Panzeri S (2013) Modelling and analysis of local field potentials for studying the function of cortical circuits. *Nat Rev Neurosci* 14:770–785. [CrossRef Medline](#)
- Freiwald WA, Tsao DY (2010) Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* 330:845–851. [CrossRef Medline](#)
- Ghazanfar AA, Schroeder CE (2006) Is neocortex essentially multisensory? *Trends Cogn Sci* 10:278–285. [CrossRef Medline](#)
- Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK (2005) Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *J Neurosci* 25:5004–5012. [CrossRef Medline](#)
- Gross CG, Rocha-Miranda CE, Bender DB (1972) Visual properties of neurons in inferotemporal cortex of the macaque. *J Neurophysiol* 35:96–111. [Medline](#)
- Hackett TA, Preuss TM, Kaas JH (2001) Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *J Comp Neurol* 441:197–222. [CrossRef Medline](#)
- Hackett TA, de la Mothe LA, Camalier CR, Falchier A, Lakatos P, Kajikawa Y, Schroeder CE (2014) Feedforward and feedback projections of caudal belt and parabelt areas of auditory cortex: refining the hierarchical model. *Front Neurosci* 8:72. [CrossRef Medline](#)
- Haegens S, Barczak A, Musacchia G, Lipton ML, Mehta AD, Lakatos P, Schroeder CE (2015) Laminar profile and physiology of the α rhythm in primary visual, auditory and somatosensory regions of neocortex. *J Neurosci* 35:14341–14352. [CrossRef Medline](#)
- Hoffman KL, Ghazanfar AA, Gauthier I, Logothetis NK (2008) Category-specific responses to faces and objects in primate auditory cortex. *Front Syst Neurosci* 1:2. [CrossRef Medline](#)
- Hunt MJ, Falinska M, Łęski S, Wójcik DK, Kasicki S (2011) Differential effects produced by ketamine on oscillatory activity recorded in the rat hippocampus, dorsal striatum and nucleus accumbens. *J Psychopharmacol* 25:808–821. [CrossRef Medline](#)
- Jellema T, Brunia CH, Wadman WJ (2004) Sequential activation of microcircuits underlying somatosensory-evoked potentials in rat neocortex. *Neuroscience* 129:283–295. [CrossRef Medline](#)
- Kajikawa Y, Schroeder CE (2011) How local is the local field potential? *Neuron* 72:847–858. [CrossRef Medline](#)
- Kajikawa Y, Schroeder CE (2015) Generation of field potentials and modulation of their dynamics through volume integration of cortical activity. *J Neurophysiol* 113:339–351. [CrossRef Medline](#)
- Kajikawa Y, Frey S, Ross D, Falchier A, Hackett TA, Schroeder CE (2015) Auditory properties in the parabelt regions of the superior temporal gyrus in the awake macaque monkey: an initial survey. *J Neurosci* 35:4140–4150. [CrossRef Medline](#)
- Kayser C, Petkov CI, Augath M, Logothetis NK (2007a) Functional imaging reveals visual modulation of specific fields in auditory cortex. *J Neurosci* 27:1824–1835. [CrossRef Medline](#)
- Kayser C, Petkov CI, Logothetis NK (2007b) Tuning to sound frequency in auditory field potentials. *J Neurophysiol* 98:1806–1809. [CrossRef Medline](#)
- Kayser C, Petkov CI, Logothetis NK (2008) Visual modulation of neurons in auditory cortex. *Cereb Cortex* 18:1560–1574. [CrossRef Medline](#)
- Kayser C, Logothetis NK, Panzeri S (2010) Visual enhancement of the information representation. *Curr Biol* 20:19–24. [CrossRef Medline](#)
- Kral A, Schröder JH, Klinke R, Engel AK (2003) Absence of cross-modal reorganization in the primary auditory cortex of congenitally deaf cats. *Exp Brain Res* 153:605–613. [CrossRef Medline](#)
- Lakatos P, O’Connell MN, Barczak A, Mills A, Javitt DC, Schroeder CE (2009) The leading sense: supramodal control of neurophysiological context by attention. *Neuron* 64:419–430. [CrossRef Medline](#)
- Legatt AD, Arezzo J, Vaughan HG Jr (1980) Averaged multiple unit activity as an estimate of phasic changes in local neuronal activity: effects of volume-conducted potentials. *J Neurosci Methods* 2:203–217. [CrossRef Medline](#)
- Lipton ML, Liszewski MC, O’Connell MN, Mills A, Smiley JF, Branch CA, Isler JR, Schroeder CE (2010) Interactions within the hand representation in primary somatosensory cortex of primates. *J Neurosci* 30:15895–15903. [CrossRef Medline](#)
- Matsuo T, Kawasaki K, Osada T, Sawahata H, Suzuki T, Shibata M, Miyakawa N, Nakahara K, Iijima A, Sato N, Kawai K, Saito N, Hasegawa I (2011) Intranasal electrocorticography in macaque monkeys with minimally invasive neurosurgical protocols. *Front Syst Neurosci* 5:34. [CrossRef Medline](#)
- McLaughlin DF, Juliano SL (2005) Disruption of layer 4 development alters laminar processing in ferret somatosensory cortex. *Cereb Cortex* 15:1791–1803. [CrossRef Medline](#)
- Meyer T, Olson CR (2011) Statistical learning of visual transitions in mon-

- key inferotemporal cortex. *Proc Natl Acad Sci U S A* 108:19401–19406. [CrossRef Medline](#)
- Müller-Preuss P, Mitzdorf U (1984) Functional anatomy of the inferior colliculus and the auditory cortex: current source density analyses of click-evoked potentials. *Hear Res* 16:133–142. [CrossRef Medline](#)
- Nunez PL, Srinivasan R (2006) *Electric fields of the brain*. New York: Oxford UP.
- Paxinos G, Huang XF, Petrides M, Toga AW (2008) *The rhesus monkey brain in stereotaxic coordinates*, Ed 2. San Diego: Academic.
- Perrett DI, Rolls ET, Caan W (1982) Visual neurons responsive to faces in the monkey temporal cortex. *Exp Brain Res* 47:329–342. [Medline](#)
- Pettersen KH, Devor A, Ulbert I, Dale AM, Einevoll GT (2006) Current-source density estimation based on inversion of electrostatic forward solution: effects of finite extent neuronal activity and conductivity. *J Neurosci Methods* 154:116–133. [CrossRef Medline](#)
- Pinsk MA, DeSimone K, Moore T, Gross CG, Kastner S (2005) Representations of faces and body parts in macaque temporal cortex: a functional MRI study. *Proc Natl Acad Sci U S A* 102:6996–7001. [CrossRef Medline](#)
- Recanzone GH, Guard DC, Phan ML (2000) Frequency and intensity response properties of single neurons in the auditory cortex of behaving macaque monkey. *J Neurophysiol* 83:2315–2331. [Medline](#)
- Riehle A, Wirtsohn S, Grün S, Brochier T (2013) Mapping the spatio-temporal structure of motor cortical LFP and spiking activities during reach-to-grasp movements. *Front Neural Circuits* 7:48. [CrossRef Medline](#)
- Saleem K, Logothetis N (2012) A combined MRI and histology atlas of the Rhesus monkey brain in stereotaxic coordinates, Ed 2. San Diego: Academic.
- Saleem KS, Tanaka K, Rockland KS (1993) Specific and columnar projection from area TEO to TE in the macaque inferotemporal cortex. *Cereb Cortex* 3:454–464. [CrossRef Medline](#)
- Schroeder CE, Foxe JJ (2002) The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Brain Res Cogn Brain Res* 14:187–198. [CrossRef Medline](#)
- Schroeder CE, Tenke CE, Givre SJ, Arezzo JC, Vaughan HG Jr (1991) Striate cortical contribution to the surface-recorded pattern-reversal VEP in the alert monkey. *Vision Res* 31:1143–1157. [CrossRef Medline](#)
- Schroeder CE, Steinschneider M, Javitt DC, Tenke CE, Givre SJ, Mehta AD, Simpson GV, Arezzo JC, Vaughan HG Jr (1995) Localization of ERP generators and identification of underlying neural processes. *Electroencephalogr Clin Neurophysiol Suppl* 44:55–75. [Medline](#)
- Schroeder CE, Mehta AD, Givre SJ (1998) A spatiotemporal profile of visual system activation revealed by current source density analysis in the awake macaque. *Cereb Cortex* 8:575–592. [CrossRef Medline](#)
- Schroeder CE, Lakatos P, Kajikawa Y, Partan S, Puce A (2008) Neuronal oscillations and visual amplification of speech. *Trends Cogn Sci* 12:106–113. [CrossRef Medline](#)
- Seltzer B, Pandya DN (1989) Intrinsic connections and architectonics of the superior temporal sulcus in the rhesus monkey. *J Comp Neurol* 290:451–471. [CrossRef Medline](#)
- Seltzer B, Cola MG, Gutierrez C, Masee M, Weldon C, Cusick CG (1996) Overlapping and nonoverlapping cortical projections to cortex of the superior temporal sulcus in the rhesus monkey: double anterograde tracer studies. *J Comp Neurol* 370:173–190. [CrossRef Medline](#)
- Steinschneider M, Tenke CE, Schroeder CE, Javitt DC, Simpson GV, Arezzo JC, Vaughan HG Jr (1992) Cellular generators of the cortical auditory evoked potential initial component. *Electroencephalogr Clin Neurophysiol* 84:196–200. [CrossRef Medline](#)
- Sugase Y, Yamane S, Ueno S, Kawano K (1999) Global and fine information coded by single neurons in the temporal visual cortex. *Nature* 400:869–873. [CrossRef Medline](#)
- Takeuchi D, Hirabayashi T, Tamura K, Miyashita Y (2011) Reversal of interlaminar signal between sensory and memory processing in monkey temporal cortex. *Science* 331:1443–1447. [CrossRef Medline](#)
- Tanaka K, Saito H, Fukada Y, Moriya M (1991) Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J Neurophysiol* 66:170–189. [Medline](#)
- Tsao DY, Moeller S, Freiwald WA (2008) Comparing face patch systems in macaques and humans. *Proc Natl Acad Sci U S A* 105:19514–19519. [CrossRef Medline](#)
- Turesson HK, Logothetis NK, Hoffman KL (2012) Category-selective phase coding in the superior temporal sulcus. *Proc Natl Acad Sci U S A* 109:19438–19443. [CrossRef Medline](#)
- Ungerleider LG, Galkin TW, Desimone R, Gattass R (2008) Cortical connections of area V4 in the macaque. *Cereb Cortex* 18:477–499. [CrossRef Medline](#)
- van Kerkoerle T, Self MW, Roelfsema PR (2017) Layer specificity in the effects of attention and working memory on activity in primary visual cortex. *Nat Commun* 8:13804. [CrossRef Medline](#)
- Woloszyn L, Sheinberg DL (2009) Neural dynamics in inferior temporal cortex during a visual working memory task. *J Neurosci* 29:5494–5507. [CrossRef Medline](#)
- Woodman GF, Kang MS, Rossi AF, Schall JD (2007) Nonhuman primate event-related potentials indexing covert shifts of attention. *Proc Natl Acad Sci U S A* 104:15111–15116. [CrossRef Medline](#)