

AUDITORY-VISUAL PROCESSING REPRESENTED IN THE HUMAN SUPERIOR TEMPORAL GYRUS

R. A. REALE,^{a,b,*} G. A. CALVERT,^{c,d} T. THESEN,^c
R. L. JENISON,^b H. KAWASAKI,^a H. OYA,^a
M. A. HOWARD^a AND J. F. BRUGGE^{a,b}

^aDepartment of Neurosurgery, University of Iowa, Iowa City, IA 52242, USA

^bDepartment of Psychology, University of Wisconsin, 1202 West Johnson Street, Madison, WI 53706, USA

^cUniversity Laboratory of Physiology, University of Oxford, Oxford, UK

^dDepartment of Psychology, University of Bath, Bath, UK

Abstract—In natural face-to-face communication, speech perception utilizes both auditory and visual information. We described previously an acoustically responsive area on the posterior lateral surface of the superior temporal gyrus (field PLST) that is distinguishable on physiological grounds from other auditory fields located within the superior temporal plane. Considering the empirical findings in humans and non-human primates of cortical locations responsive to heard sounds and/or seen sound-sources, we reasoned that area PLST would also contain neural signals reflecting audiovisual speech interactions. To test this hypothesis, event related potentials (ERPs) were recorded from area PLST using chronically implanted multi-contact subdural surface-recording electrodes in patient-subjects undergoing diagnosis and treatment of medically intractable epilepsy, and cortical ERP maps were acquired during five contrasting auditory, visual and bimodal speech conditions. Stimulus conditions included consonant–vowel (CV) syllable sounds alone, silent seen speech or CV sounds paired with a female face articulating matched or mismatched syllables. Data were analyzed using a MANOVA framework, with the results from planned comparisons used to construct cortical significance maps. Our findings indicate that evoked responses recorded from area PLST to auditory speech stimuli are influenced significantly by the addition of visual images of the moving lower face and lips, either articulating the audible syllable or carrying out a meaningless (gurning) motion. The area of cortex exhibiting this audiovisual influence was demonstrably greater in the speech-dominant hemisphere. © 2006 IBRO. Published by Elsevier Ltd. All rights reserved.

Key words: temporal lobe, audiovisual processing, speech.

*Correspondence to: R. A. Reale, Department of Psychology, University of Wisconsin, 1202 West Johnson Street, Madison, WI 53706, USA. Tel: +1-608-263-5851; fax: +1-608-263-5929. E-mail address: rreale@wisc.edu (R. A. Reale).

Abbreviations: AI, primary auditory cortex; AV, audiovisual; AW, analysis window; CV, consonant–vowel; ECoG, electrocorticogram; EEG, electroencephalogram; ERPs, event-related potentials; fMRI, functional magnetic resonance imaging; FWE, family-wise (type-I) error rate; MANOVA, multivariate analysis of variance; MEG, magnetoencephalography; MTG, middle temporal gyrus; PCA, principal component analysis; PET, positron emission tomography; PLST, posterior lateral superior temporal gyrus; SF, Sylvian fissure; STG, superior temporal gyrus; STS, superior temporal sulcus; VW, Van der Waerden.

0306-4522/07\$30.00+0.00 © 2006 IBRO. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuroscience.2006.11.036

In natural face-to-face communication, speech perception engages neural processes that integrate acoustic and visual information. Under these conditions listeners naturally and effortlessly create a unified and coherent percept using complementary information obtained through hearing the speaker's voice and seeing the articulatory movements of a speaker's face. As a result, speech perception is enhanced, presumably using neural mechanisms of bimodal interaction (Fowler, 2004; Massaro, 1998, 2004; Munhall and Vatikiotis-Bateson, 2004). The impact of one modality upon the perception of the other during audiovisual (AV) speech processing is especially important in noisy or reverberant environments, or when hearing is otherwise impaired (Campbell and Dodd, 1980; Dodd, 1977; Grant and Seitz, 2000; Sumbly and Pollack, 1954; Summerfield, 1987, 1992).

The integrative mechanisms of the perceptual system that bind auditory and visual modalities are complex and in the broadest sense involve neural circuitry distributed across the frontal, parietal and temporal lobes (Calvert and Lewis, 2004; Hall et al., 2005; Kaas and Collins, 2004) as well as subcortical structures long known to exhibit multi-sensory interactions (Stein and Meredith, 1993; Wallace et al., 2004). Temporal lobe cortex within or close to the banks of the superior temporal sulcus (STS) has been identified as a region of multimodal integration and interactions in both humans (Calvert and Lewis, 2004; Wright et al., 2003) and non-human primates (Barraclough et al., 2005; Baylis et al., 1987; Benevento et al., 1977; Bruce et al., 1981; Desimone and Gross, 1979; Hikosaka et al., 1988). Hemodynamic and magnetoencephalography (MEG) studies in humans also suggest that AV speech interactions occur within even more extensive areas of the temporal lobe including cortex of the superior temporal gyrus (STG) and superior temporal plane (Calvert and Lewis, 2004; Karnath, 2001; Kuriki et al., 1995; Raj and Jousmaki, 2004). These latter cortices comprise the most consistently identified locations for secondary (associative) and primary auditory cortical fields (Binder et al., 1997; Creutzfeldt et al., 1989; Galaburda and Sanides, 1980; Hackett et al., 2001; Howard et al., 2000; Liegeois-Chauvel et al., 1991; Rivier and Clarke, 1997; Sweet et al., 2005; Wallace et al., 2002). Studies of scalp-recorded event-related potentials (ERPs) in human have consistently implicated auditory cortex as a site for both AV and audio-tactile multimodal integration (Besle et al., 2004; Foxe et al., 2000; Giard and Peronnet, 1999; Molholm et al., 2002; Murray et al., 2005; van Wassenhove et al., 2005). These implications are well supported by hemodynamic evidence in human (Calvert et al., 1999, 2000; Foxe et al., 2002;

Pekkola et al., 2005; van Atteveldt et al., 2004) and non-human primate (Kayser et al., 2005). Direct intracranial recordings from macaque auditory cortex revealed that audio-tactile (Fu et al., 2003; Schroeder et al., 2001; Schroeder and Foxe, 2002) and AV (Ghazanfar et al., 2005) multimodal convergence occurs in auditory cortices posterior to the primary auditory area including at least one secondary field (Schroeder et al., 2003). Taken together these findings have promoted a framework for multisensory processing that emphasizes convergence and integration at the earliest stages of auditory cortical processing (for review see Ghazanfar and Schroeder, 2006; Schroeder et al., 2003).

Auditory cortex of the temporal lobe of human is composed of multiple fields although, with the exception of the primary field (primary auditory cortex, AI), there is still no agreement on the number of fields or their spatial locations (reviewed by Hackett, 2003). These fields are thought to be organized in a three-tier, core–belt–parabelt, hierarchical processing system, similar to that proposed for monkey (Hackett et al., 1998a; Kaas and Hackett, 2000; Rauschecker and Tian, 2000; Sweet et al., 2005; Wessinger et al., 2001). Only the core primary field, AI, is considered to be homologous between monkey and human (Hackett et al., 2001; Hackett, 2003). Thus, for the present, studies of functional localization outside of the auditory core in humans must be carried out on humans. Noninvasive imaging methods (functional magnetic resonance imaging (fMRI), positron emission tomography (PET), MEG, electroencephalogram (EEG)) provide powerful approaches to studies of functional localization in the human brain, and these have been applied to cortical localization of AV interaction patterns (fMRI: Callan et al., 2003, 2004; Calvert et al., 1997, 1999, 2000; Calvert, 2001; MacSweeney et al., 2002; Pekkola et al., 2005; van Atteveldt et al., 2004; MEG: Mottonen et al., 2002; Sams et al., 1991, and EEG: Callan et al., 2001; Klucharev et al., 2003). Taken together these studies suggest that multisensory interactions occur within portions of the STS, middle and superior temporal gyri as well as cortical areas traditionally considered to overlap with auditory cortex. Which of the multiple auditory fields on STG represent these interactions is, however, a question not answered by noninvasive approaches because functional imaging studies have not been combined with the locations of cytoarchitecturally or electrophysiologically identified fields. The question can be addressed by systematic mapping of stimulus-evoked activity recorded directly from the STG. Under these conditions relatively precise functional localization can be achieved in individual human subjects. Heretofore, however, no direct recordings in humans from any physiologically identified auditory cortical field have been shown to exhibit multisensory interactions.

We described previously an acoustically responsive area on the posterior lateral surface of the STG (field posterior lateral superior temporal gyrus (PLST)) that is distinguishable on physiological grounds from auditory fields located on Heschl's gyrus (HG) within the superior temporal plane (Howard et al., 2000). Considering the

empirical findings in humans and non-human primates of cortical locations responsive to heard sounds and/or seen sound-sources, we reasoned that area PLST would contain neural signals reflecting AV speech interactions. To test this hypothesis we recorded directly from PLST and surrounding perisylvian cortex and compared ERPs obtained there to auditory, visual and AV stimuli. Recordings were made using multi-contact subdural surface-recording electrodes chronically implanted in patients undergoing diagnosis and treatment of medically intractable epilepsy. Five stimuli were created for these experiments: 1) an audible consonant–vowel (CV) syllable alone, 2) an articulating female face alone, 3) an articulating female face paired with this female's natural audible utterance of a congruent CV syllable, 4) an audible CV syllable paired with meaningless lip movement, and 5) an audible CV syllable paired with lip movement of a different syllable.

The ERP was considered a multivariate observation in which the ordered sequence of measured voltages defined a *response vector* that depended upon three experimental factors: Stimulus Type, Electrode Site, and Analysis Window (AW). We used multivariate analysis of variance (MANOVA) to test inferences based on contrasts between these factors and constructed maps showing the cortical locations of significant effects. We interpret significant contrasts between the bimodal response to combined heard and seen speech (AV speech) and the unimodal response to either modality presented alone as reflecting an AV interaction. Our findings indicate that within area PLST evoked responses to auditory speech stimuli are influenced significantly by the addition of visual images of the moving lower face and lips, either articulating the audible syllable or carrying out a meaningless (gurning) motion. Moreover, this AV influence was demonstrably more extensive on the speech-dominant hemisphere.

EXPERIMENTAL PROCEDURES

Human subjects

The eight subjects (six male, two female) in this study were patients undergoing diagnosis and, later, surgical treatment for medically intractable epilepsy. They ranged in age from 18 to 56 years. As part of their clinical treatment plan multi-contact recording grids were implanted over the perisylvian cortex and for the following 2 weeks their electrocorticogram (ECoG) was monitored continuously for seizure activity. The recording grid was on the left hemisphere (L) of four subjects and the right (R) of four. An intra-arterial amobarbital procedure (WADA test) was carried out on seven subjects, and the results indicated that the left cerebral hemisphere of each was dominant for speech. One subject (L122) did not undergo WADA testing, and hence cerebral dominance for speech in this subject is uncertain. Research recordings were obtained from the grid electrodes while the subjects were reclining comfortably in a bed or sitting upright in a chair. Speech tokens were delivered through calibrated insert earphones while visual images were presented on a video monitor positioned directly in front of the subject. Recording sessions were carried out either in the epilepsy ward or in our specially designed and equipped human electrophysiological recording facility. Informed consent was obtained after the nature and possible consequences of the studies were explained to the patient. All study protocols were approved by the University of Iowa Institutional Review Board.

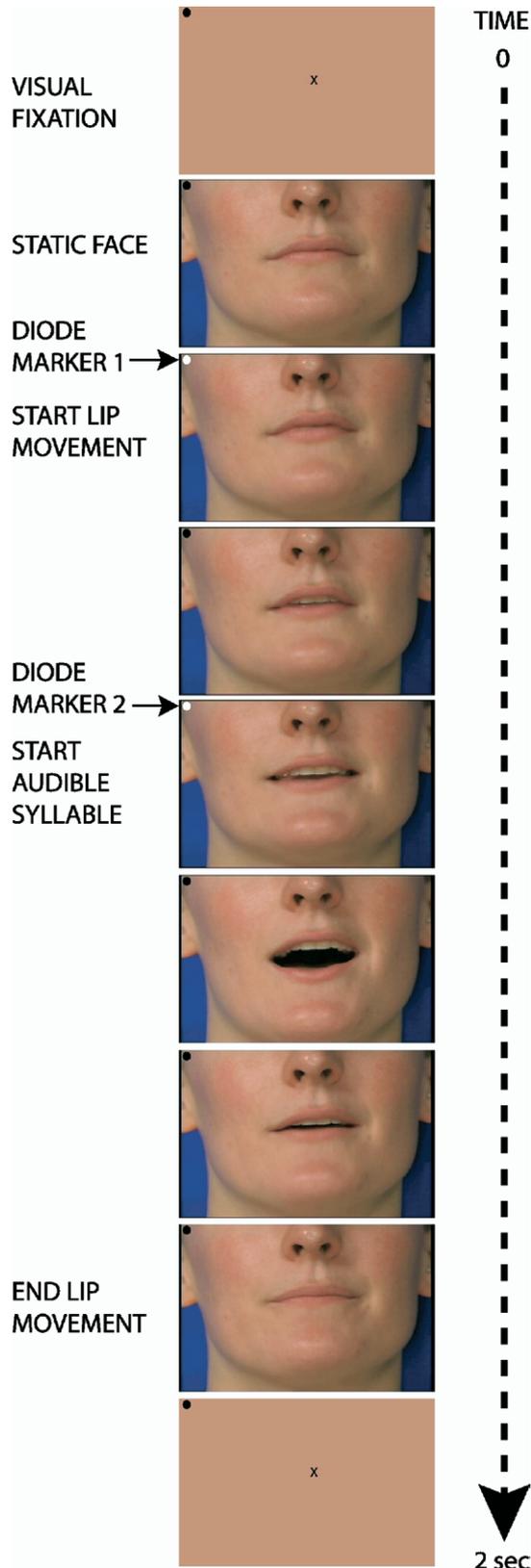


Fig. 1. Sequence of sample frames from a video clip (AVI format) used to present AV speech. The sequence begins and ends with a video screen showing a centrally placed black visual fixation point on a

Subjects did not incur additional risk by participating in these studies. Presurgical audiometric testing showed that for all subjects pure-tone thresholds (500 Hz–4000 Hz) and language comprehension were within normal limits.

No cortical abnormalities were identified in the MRI of two subjects (R104, L118). Mesial temporal sclerosis was identified in three subjects (R129, L100, L106). Subject L122 had regions of cortical dysplasia involving the middle and posterior portions of the superior and middle temporal gyri. Recordings were obtained from two subjects (R98 and R127) with lesions of cortical areas that are known to be anatomically and functionally related to the posterior STG. Subject R98, an 45-year old male, had experienced simple and complex partial seizures from the age of 13. He had undergone a partial right temporal lobectomy on October 23, 1996. At that time the right hippocampus and inferior and middle temporal gyri were removed leaving intact the STG, including the cortex that comprises the ventral bank of the STG; the middle temporal gyrus (MTG) cortex lining the ventral surface of the STS was removed completely. Hence, the cortex lining the ventral surface of the STS that is considered a major site of visual input (Seltzer and Pandya, 1978) could not contribute to the ERPs recorded on the STG when recordings were made nearly 8 years later. The boundaries of the lesion were identified by analysis of serial MRI sections. Subject R127 had, in 1979, sustained a right frontal head injury. He subsequently developed meningitis and CSF rhinorrhea. In 1980 he underwent surgical treatment of a right orbital abscess and reconstruction of the orbit. MRI revealed a right frontal encephalomalacia and large porencephalic cyst. He began having seizures in 1981. PET imaging revealed absence of metabolism in the region of the right frontal lobe injury and reduced metabolism throughout the right temporal lobe. This subject provided us the rare opportunity to examine evoked activity in STG in the absence of normal cortico-cortical input from the frontal lobe.

Stimulus construction

Audio, visual and AV stimuli were derived from videotaped segments of the lower face, including the mouth, of a female speaker articulating three syllables (/da/, /gi/, /tu/) or carrying out meaningless lip movements (gurning). We chose these syllables because we were especially interested in probing the earliest cross-modal influences of AV speech perception using non-semantic phonetic stimuli whose integration is likely to occur at a pre-lexical, phonetic categorization stage (MacSweeney et al., 2001; Summerfield, 1991; Summerfield and McGrath, 1984). They were also shown to be easily discriminated, both acoustically and visually, in two companion studies (Calvert et al., 2005; Thesen et al., 2005) that served as complementary approaches to this investigation of AV interactions. The image was restricted to the lower face because we wished to focus our attention on the role played by the articulators in AV speech (Summerfield, 1992) and to minimize the influence of possible confounding factors such as gaze and facial identity (Campbell et al., 1986, 2001). Fig. 1 illustrates nine sequential video frames from a series of 62 frames (lasting about 2000 ms) viewed by the subject in which the syllable /da/ was uttered. Natural AV speech is an ecologically valid stimulus and is characterized by the temporal precedence of visual speech, as the movement of the facial articulators usually precedes the onset of

background whose luminance and color were matched to subsequent presentations of the subjects lower face. The static face image then appeared and remained for 734 ms before the lips moved (diode marker 1) and nominally 1167 ms before the start of the audible syllable (diode marker 2). Diode markers are not visible to the subject. Onset of the acoustic syllable was derived from a digitized version of the sound waveform delivered to the earphones. The entire sequence lasted about 2 s and ended when the lips stopped moving and the original video screen appeared showing the black visual fixation point.

Table 1. Auditory and visual components and response vectors for five stimulus types

Stimulus type	Auditory signal	Visual image	Response vectors
A_{da}	/da/	None	\vec{A}_{da}
V_{da}	None	Natural lip movement for /da/	\vec{V}_{da}
$A_{da}V_{da}$	/da/	Natural lip movement for /da/	$\vec{A}_{da} + \vec{V}_{da} + \vec{INT}$
$A_{gi}V_{da}$	/gi/	Natural lip movement for /da/	$\vec{A}_{gi} + \vec{V}_{da} + \vec{INT}$
$A_{da}V_{gurn}$	/da/	Gurning	$\vec{A}_{da} + \vec{V}_{gurn} + \vec{INT}$

the acoustic stimulus by tens to hundreds of milliseconds. The series began with a video screen showing a centrally-placed black visual fixation point on a background whose luminance and color were matched to the subsequent presentations of a human lower face. The face appeared abruptly, and shortly thereafter the lips began to move followed by the start of the vocal utterance. The entire AV token ended with the appearance of the original color-matched uniform background screen.

Table 1 shows the auditory and visual components making up the five *stimulus types* for which the resulting ECoG data were analyzed. One of the AV speech tokens was congruent ($A_{da}V_{da}$), i.e. the visually articulated syllable was the same as the auditory syllable. Audio-alone (A_{da}) and visual-alone (V_{da}) tokens were produced simply by eliminating the respective visual or the audio component from the video clip. In this framework, an AV-stimulus response ($\vec{A}_{da}V_{da}$) was considered to be the summation of response vectors \vec{A}_{da} and \vec{V}_{da} elicited by the auditory and visual tokens presented separately, plus an *interaction vector* (\vec{INT}). As \vec{INT} is, by definition, elicited only by the bimodal stimulus, it is not accounted for by linear addition of unimodal responses (see also Giard and Peronnet, 1999). In order to test whether AV interactions required congruent (naturally paired) components, two incongruent AV speech tokens were constructed by pairing articulatory lip movement associated with the syllable /da/ with the audible signal of the syllable /gi/ ($A_{gi}V_{da}$), and by pairing the audible syllable /da/ with meaningless, closed-mouth movements having no relationship to any syllable utterance ($A_{da}V_{gurn}$), a visual condition known as gurning (Campbell et al., 2001). Stimulus construction was aided by the use of a commercially available digital video editing software (Premiere V6.0, Adobe Systems Inc., San Jose, CA, USA). The 62 video frames that made up an AV token were played out at 29.97 fps, with (stereo) audio signals digitized at 44.1 kHz and 16 bit resolution (digital AVI format). At the start of a video sequence a blank screen appeared with a black cross in the center. The subject was instructed to fixate on this point and then to watch carefully the subsequent facial images. We did not monitor the subject's eye movements during this series of subjects. In more recent studies using an eye-tracking apparatus, a subject's eyes commonly scanned the presentations of the human lower face during data collection.

Ten identical repetitions of each stimulus type were presented in random order from a single AVI multimedia file together with 10 trials of AV speech that utilized the articulation and utterance of the syllable /tu/. The subject was instructed to press a button whenever /tu/ was detected. This action served only to maintain the subject's vigilance during the showing of the video clip. Electrophysiological data from these 10 /tu/ trials were discarded and not analyzed. Typically, four AVI movie files (total=40 trials per token in Table 1), each with unique randomization, were available for presentation. We randomized the stimulus events in an attempt

to further reduce uncontrolled non-stationary influences (e.g. alertness, arousal).

Acoustic calibration and stimulus presentation

Prior to surgery each subject was custom fitted in the hearing aid clinic of the Univ. Iowa Department of Otolaryngology with ear molds of the kind commonly worn by hearing aid users. Acoustic stimuli were delivered binaurally over miniature earphones (ER4B, Etymotic Research, Elk Grove Village, IL, USA) integrated into each ear mold. An intensity level was chosen that was comfortable for the subject and that permitted easy discrimination of the naturally spoken syllables. Provision was also made for a probe microphone to be inserted through a second port on the ear mold for the purpose of acoustic calibration. Acoustical calibration was carried out in six of the eight subjects by presenting maximum length sequences (Golay codes, Zhou et al., 1992) and recording their responses through a clinical probe microphone (ER-7C: Etymotic Research) system. Subsequently, in these six subjects equalizer filters were constructed for sound-path channels to the left and right ears, and these filters were used to compensate the speech tokens for each channel. For the remaining two subjects, no attempt was made to compensate for the small frequency dependent differences in sound pressure level inherent in the output of the earphones. Real Player® (RealNetworks Inc., Seattle, WA, USA) was used for playback of digital AVI media files at a comfortable sound level (typically 35–40 dB above threshold) from a dedicated PC platform with a standard flat-screen CRT monitor.

Electrophysiological recording

Details of electrode implantation and data acquisition can be found in an earlier paper (Howard et al., 2000), and only a brief description will be given here. In each subject an array of platinum–iridium disc electrodes (1.6 mm diameter, 4–5 mm inter-electrode distance) embedded in a silicon membrane was implanted on the pial surface over the perisylvian regions of the temporal and parietal lobes, including the posterolateral STG where previously we had identified an auditory field (PLST, Howard et al., 2000). ECoG data were acquired (2 ksamples/s; bandpass 1–1000 Hz) simultaneously from the 64 surface-electrode recording grids implanted on the left hemisphere in three subjects (L106, L100, L122) and the right hemisphere in four subjects (R98, R104, R127, R129). In one subject (L118), 22 contacts were found to be defective and, hence, simultaneous recording was obtained from only 42 sites. MRI and intraoperative photographs aided the reconstruction of the location of the electrode grid with respect to gyral landmarks.

The number and distribution of active cortical sites recorded varied from one subject to the next, as the placement of the electrode grid depended entirely on clinical considerations. In each of the present subjects, the multi-contact surface-electrode recording grid had been largely confined to the posterolateral STG. This limitation precluded systematic investigations of more anterior temporal cortex. We did not study sites outside of the electrode arrays illustrated in this paper, as these were the only grids of this type implanted. Four-contact strip electrodes were placed beneath the inferior temporal gyrus, but these recordings were used exclusively for clinical purposes.

ERPs obtained in response to stimuli that contained an audible syllable were referenced in time to the onset of the syllable. This syllable onset was derived from a digitized copy of the audio waveform delivered to the subject's earphone. For the visual-alone stimulus type, ERPs were timed with respect to the video frame that correlated with the onset of the (removed) audible syllable (nominally 432 ms after the onset of lip movement). This fiducial time stamp was marked by the appearance of a small

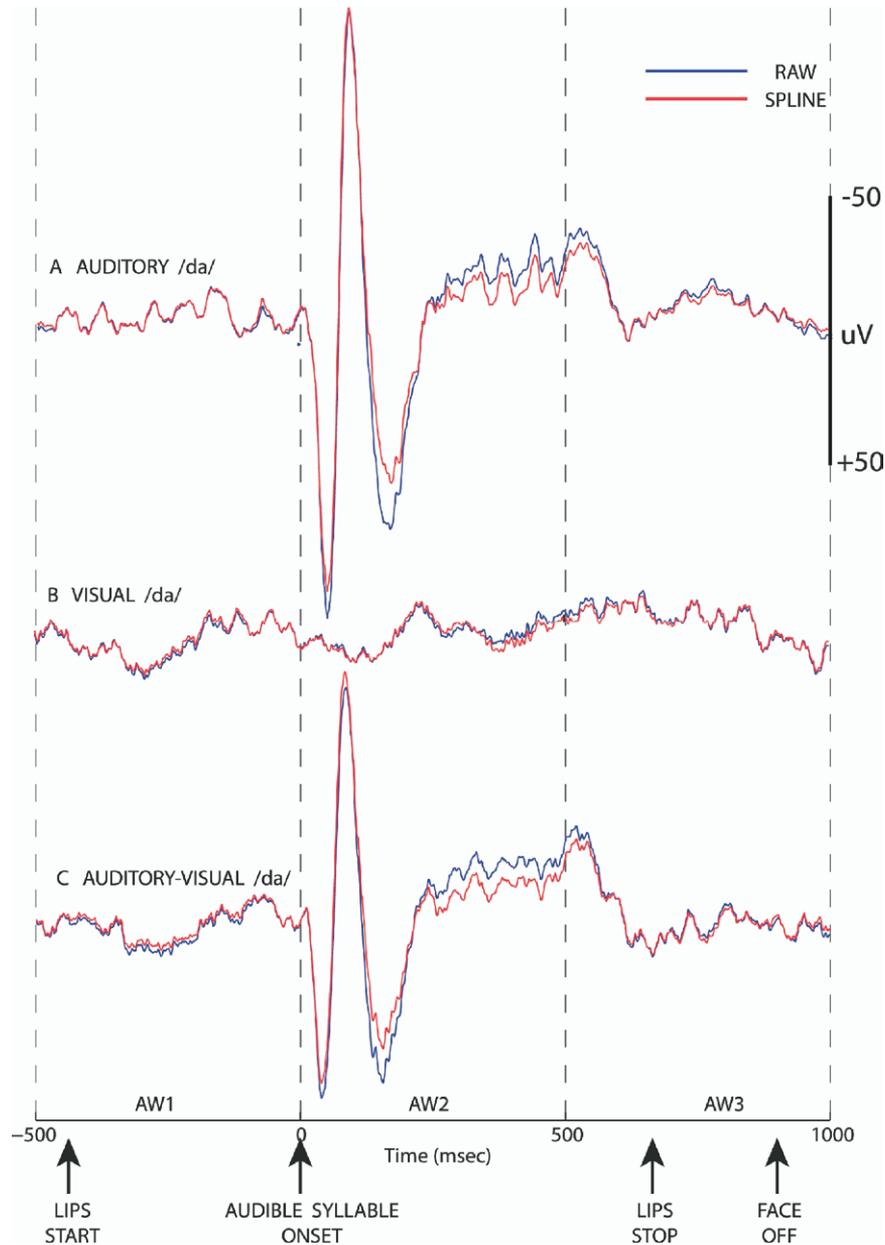


Fig. 2. Average ERP waveforms evoked by unimodal (Auditory or Visual) and bimodal (AV) speech and recorded at the site of maximal voltage in subject L106. Measured voltages (blue) and their spline-fit counterparts (red) are shown for each stimulus type. Vertical dashed lines mark the temporal boundaries of the three 500 ms AWs included in the MANOVA analysis. A common ordinate scale is used for all waveforms. Negative voltage plotted in the upward direction.

white circle (unseen by the subject) in the upper left corner of the screen (see Fig. 1). The appearance of this white circle was detected by a diode that emitted a TTL signal, that was digitized on still another A/D channel. These multiple time markers were needed to synchronize responses with stimulus constructs, as the operating system of a PC could interrupt playback at unpredictable times.

Research recording usually began 2 days after electrode implantation. At this time, an image of the recording grid was superimposed upon a preoperative MRI of the lateral surface of the cerebral hemisphere using gyral landmarks derived from photographs taken during implantation surgery. Later, when the recording grid was removed, the grid position was verified and, if necessary, its representation on

the postoperative MRI adjusted appropriately. Typically, in the initial recording sessions, we obtained ERP maps using only audible stimuli including clicks, noise bursts, tone bursts and syllables. This preliminary recording served to acquaint the subject with the recording sessions, to identify technical difficulties that might have arisen during early surgical recovery, to test the responsiveness of the cortex beneath the grid, and to map the location and boundaries of area PLST. Duration of daily recording times was controlled by the subject's consent. In practice, this limitation precluded a parametric exploration of stimulus variables (e.g. interval between, or intensity of, visual and auditory stimulus components) that are commonly studied with animal models of AV interactions.

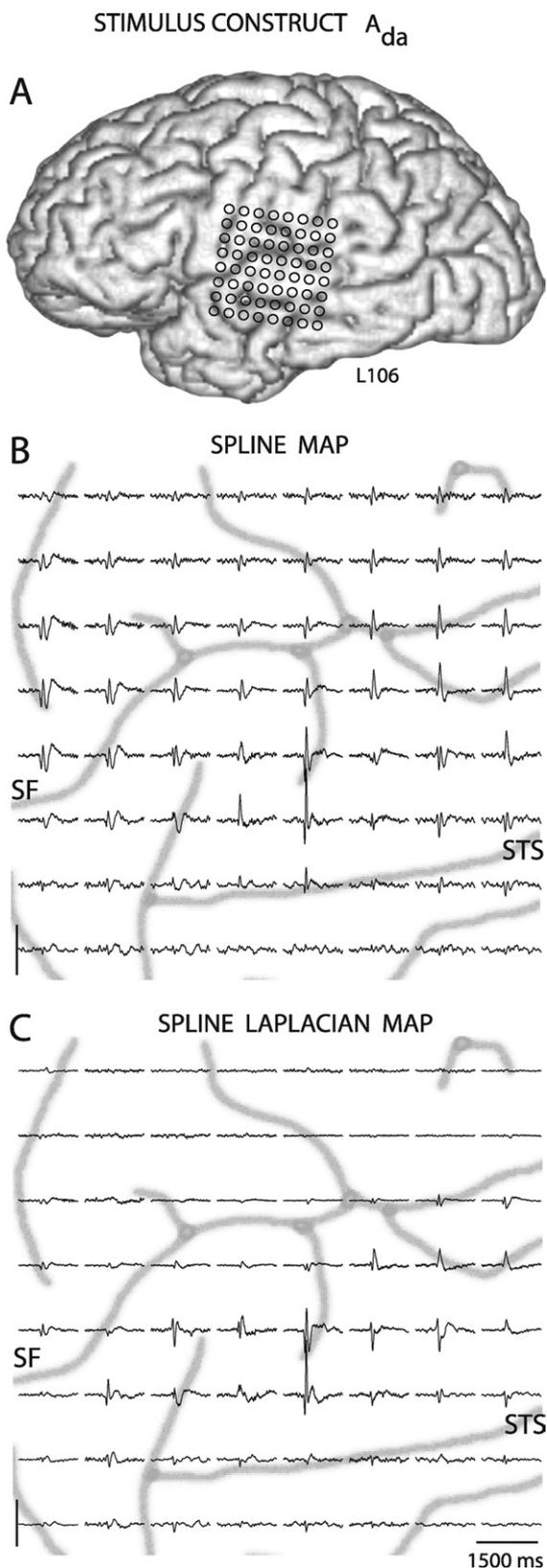


Fig. 3. Spline–Laplacian transformation. (A) Location of 64-contact recording grid overlying perisylvian cortex. (B) Average ERP at each of the original recording sites derived from spline fit. Ordinate scale

Spline–Laplacian transformation

The ERP is the result of a series of local synaptic current sinks and sources triggered by the invasion of stimulus-evoked input arriving over one or more afferent pathways. Based on evidence from single- and multi-neuron recording and current source density measurement, it is generally accepted that the ERP waveform recorded by an electrode on the brain surface reflects these physiologic events occurring mainly within the cortex in some restricted region beneath the recording electrode (Creutzfeldt and Houchin, 1984; reviewed by Arezzo et al., 1986; Vaughan and Arezzo, 1988; Mitzdorf, 1991, 1994; Steinschneider et al., 1992). Nevertheless, the spatial distribution of the potentials sampled by our electrodes was necessarily influenced by the choice of the reference electrode and the effects of spatial smearing due to volume conduction in tissue and fluid of the brain. In order to help ameliorate these influences, the distribution of the measured potential was transformed with a spatial filter using the surface Laplacian operation (Nunez and Westdorp, 1994; Nunez and Pilgreen, 1991; Nunez, 1981; Perrin et al., 1987; Law et al., 1993). The surface Laplacian is independent of the reference electrode and is proportional to the so-called current source density. It is now appreciated that the 3-D Laplacian of the scalp EEG potential is a good estimate of the spatial distribution of the dura/pial surface potentials (Nunez and Westdorp, 1994; Nunez and Srinivasan, 2006). In this application, as in ours, the Laplacian acts as a high-pass spatial filter that de-emphasizes deep brain sources and/or coherent sources distributed over large cortical areas. The nature of the Laplacian is to improve the spatial resolution of more local cortical sources. From a physiological point of view, estimation of the cortical sources underlying the dura/pial surface potential, requires adoption of models for volume conduction, and cortical current sources (e.g. dipole approximations). This report does not provide data that bear on these issues. Rather we employ the surface-Laplacian method solely as a spatial filter.

The surface Laplacian required an accurate representation of the spatial distribution of potential that is generally derived using spline interpolation (Perrin et al., 1987; Law et al., 1993). Thus the Spline–Laplacian transformation required two steps. First, the distribution of voltages sampled by the electrode array was interpolated using a two-dimensional natural cubic spline to yield a (high-resolution) continuously smooth function of potential across the two-dimensional recording grid. Since the spline and the subsequent Laplacian are analytical solutions, they can be used to interpolate potentials at any electrode location (e.g. locations of ‘bad’ electrodes) for display or statistical comparisons. Second, the surface Laplacian was calculated using an analytic solution to the second-order spatial derivatives required for the computation. Our higher-order Spline–Laplacian, with units of voltage per unit-area, was derived exclusively using analytical mathematics in which the spline coefficients were estimated in a least squares sense. This has the advantage over numerical techniques where the computational overhead to estimate a suitably dense interpolated grid becomes unnecessary. Fig. 2 shows for one subject the averaged ERP waveforms (blue) measured at one electrode site (chosen for maximal voltage excursion) within PLST for the first three stimulus constructs shown in Table 1. This is compared with the waveform derived from the spline fit (red) at this electrode site. The comparison is representative of the degree to which a measured ERP and its derived spline representation agree; for most of the waveform the two curves superimpose, with some exceptions noted around waveform peaks and valleys.

–100 to 100 μV . (C) Laplacian transformation of the spline fit used to illustrate average Spline–Laplacian waveform at each of the original recording sites. Ordinate scale –325 to 325 $\mu\text{V}/\text{cm}^2$. The SF and STS are shown in gray on the spatial maps of waveforms.

The most prominent negative and positive deflections in ERPs occur in response to those stimuli having an audible syllable component (A_{da} and $A_{da}V_{da}$) and are largely confined to the 500 ms window following syllable onset (Fig. 2 A and C). Remarkably fewer prominent deflections occur in the succeeding 500 ms window, and even fewer noticeable peaks and valleys in the 500 ms window preceding syllable onset (but during visual stimulation). These features were also typical of ERPs recorded from area PLST in response to audible-only stimuli including clicks, noise bursts, tone bursts and syllables (see also Brugge et al., 2005; Howard et al., 2000) and guided our selection of AWs used as a factor in the MANOVA analysis described below.

Fig. 3 illustrates, for the A_{da} stimulus, the effects of applying the Spline–Laplacian transformation to the ERPs measured simultaneously at the 64 electrode sites depicted on the rectangular grid overlying perisylvian cortex. The response field in Fig. 3B is illustrated using the average Spline ERPs. There, waveforms characteristic of area PLST, having clearly distinguishable positive and negative deflections and overlying STG, are flanked by similar waveforms at some sites above the Sylvian fissure (SF) and below the STS (Howard et al., 2000). Subjecting these data to the Laplacian transformation (Fig. 3C) resulted in a response field that was independent of the reference electrode and therefore emphasized local sources of current largely confined to the STG. There is a close correspondence between pre- and post-transformation shapes of ERPs at some but not all electrode sites. This is to be expected, as the Spline–Laplacian transformation reduces voltage contributions from distant sites. Furthermore, Spline–Laplacian estimates are not expected to be very accurate near the edge of an electrode grid (Nunez and Pilgreen, 1991) where the disagreement in shapes can be most marked. We attempted to reduce this edge effect during spline interpolation (not entirely successfully) by adding extra electrode sites along each edge of the recording grid (see online Supplementary Data) and requiring their voltages to be zero. In this study, analyses of differences between response fields arising from different experimental factors always employed ERPs transformed by the Spline–Laplacian.

Statistical analysis: MANOVA

The analysis of variance model we employed is one commonly used to test hypotheses concerning the effects of differences among two or more experimental factors on the dependent *univariate* response measurement. In our studies, the dependent ERP measurement was treated as a multivariate *response vector* and assumed to be sampled from a multivariate normal distribution (Donchin, 1966). In this approach, MANOVA is a suitable framework in which to test whether ERPs differ among our three experimental factors: Stimulus Type, Electrode Site and AW (Dillon and Goldstein, 1984). A three-way, repeated-measures MANOVA provided an overall test of the equality of these multivariate ERP vectors as well as tests for main effects (Stimulus: 5 levels, Electrode Sites: 42–64 levels, AWs: 3 levels) and effects due to combinations of experimental factors. Thus, our statistical analysis does not depend upon the measurement of a single deflection in the ERP (since the *response variable* is a *vector* representing the ERP waveform) nor does it depend upon a simple difference in the response variable (e.g. additive or subtractive). Rather, any systematic difference between the contrasted waveforms beyond chance is sufficient to mark an effect.

The first statistic of interest was the *omnibus combination effect*, which indicated whether differences among ERPs depended on a conjunction between levels of stimulus, electrode site, and AW. In MANOVA, when the classification has more than one factor, and omnibus tests for main effects and their combinations are significant, it is common to test (i.e. contrast) the means of each level of each factor and their combinations, adjusting the resulting *P*-values to reflect these multiple comparisons. When the experimenter's primary interest is in a set of focused (e.g. single-

Table 2. Response model and five planned contrasts used to identify electrode sites exhibiting a significant effect

Response model: $\vec{AV} = \vec{A} + \vec{V} + \vec{INT}$		
	Planned contrast	Response vectors tested
C1	$\vec{A}_{da}V_{da}$ vs. \vec{A}_{da}	$\vec{V}_{da} + \vec{INT}$
C2	$\vec{A}_{da}V_{da}$ vs. \vec{V}_{da}	$\vec{A}_{da} + \vec{INT}$
C3	$\vec{A}_{da}V_{gurn}$ vs. \vec{A}_{da}	$\vec{V}_{gurn} + \vec{INT}$
C4	$\vec{A}_{da}V_{da}$ vs. $\vec{A}_{da}V_{gurn}$	$\vec{V}_{da} - \vec{V}_{gurn}$
C5	$\vec{A}_{da}V_{da}$ vs. $\vec{A}_{gi}V_{da}$	$\vec{A}_{da} - \vec{A}_{gi}$

degree-freedom) tests one can safely ignore the omnibus results and simply construct these multiple comparison tests. We chose to be conservative on this, and introduced the omnibus test which, if found to be significant ($\alpha < 0.05$), led us to carry out five planned contrasts (Table 2) to identify those electrode sites contributing to the proposed effect. Significant (non-zero) differences that arose from using these contrasts were, depending on the comparison, interpreted to reflect an AV interaction, differences in the unimodal responses, or both.

The MANOVA procedure was preceded by a principal component analysis (PCA) in order simply to reduce the dimensionality of the ERP data vectors (Hotelling, 1933; Suter, 1970). It would not have been possible to carry out the multivariate analysis using the original ERP vectors of such high dimensionality (i.e. 500 sample times). However, there is redundancy within this temporal dimension so that linear combinations of the original sample-time variables were replaced with a set of new uncorrelated principal component scores. For each subject the input to the PCA was the corpus of ERPs comprising all trials for all stimulus types, AWs, and electrode sites, after down-sampling to 1 ksample/s. For our subjects this translated to between 45,000 and 80,000 ERPs using the three AWs described above (Fig. 2). The use of the PCA scores (i.e. weights) in the ERP vectors, rather than the original voltage measurements, does not depend upon an assumption of orthogonality but only upon the adequacy of the represented ERPs (see online Supplementary Data: Methods). A sufficient number of PCs (from 14-to-21) was retained to account for 90% of the variance, which typically represented at least an order-of-magnitude reduction (e.g. 500 to <50) in the dimension of the input vectors. In practice, we employed ERPs transformed by the Spline–Laplacian as the input vectors for the PCA computation.

Cortical significance maps

Spatial maps of average ERPs, like that shown in Fig. 3B, illustrate the mean spatio-temporal relationships expressed by the neural signals in response to a particular stimulus. In order to make comparisons among response fields corresponding to different stimuli, ERPs (transformed by the Spline–Laplacian) were analyzed using a three-way MANOVA with a doubly-multivariate repeated-measures design (SAS v9.1, SAS Institute Inc., Cary, NC, USA), as described above. The AW levels were chosen by the temporal relationship between audible syllable onset and the major deflections in an ERP (see Fig. 2). A more detailed temporal analysis is beyond the scope of this paper, and will be the subject of a subsequent article.

In all 8 subjects, the omnibus effect (Stimulus \times AW \times Electrode Site) was significant at the 0.05 level. Contrast coding was then used to test the five comparisons listed in Table 2 at each of the electrode sites. These pair-wise comparisons among

stimuli allowed for the construction of cortical significance maps by marking the location of each electrode site at which the corresponding (multiple-comparison adjusted) P -value for that contrast was <0.05 . Performing multiple comparisons required this adjustment to the raw P -values in order to control for the inflation of the family-wise (type-I) error rate (FWE). This inflation in the probability of making at least one false rejection of a null hypothesis is always of concern when the entire family of inferences is considered rather than any single family member. When the dependent measure is univariate, there are well-known post hoc tests and adjustments to control for the FWE. Such procedures are not generally available for our multivariate ERP data. In this case, however, ‘generalized Bonferroni methods’ were found to work reasonably well as multiple inference procedures. These procedures control the error rate for testing individual hypotheses, not just the overall null. The Bonferroni method guarantees strict control of the FWE when the concern is with either multiple hypotheses or simultaneous confidence intervals even when the comparisons are not necessarily independent. Since our interest was only in simultaneous tests of hypotheses, FWE was controlled and power improved by using the Hochberg-modified method (Westfall et al., 1999).

RESULTS

We described previously an area of cortex on the posterolateral STG (field PLST) that could be activated, bilaterally, by a wide range of acoustic stimuli including clicks, puretones, band-pass noise, and syllable utterances (Brugge et al., 2005; Howard et al., 2000). Fig. 4 illustrates, for the eight subjects in the current study, the average ERP waveform recorded at the site of maximal responsiveness within PLST to three stimulus types: auditory /da/ alone (A_{da} , blue), visual /da/ alone (V_{da} , green) and the congruent auditory-visual /da/ ($A_{da}V_{da}$, red). The average ERPs at this and all other sites within PLST exhibited prominent positive and negative voltage deflections in response to the acoustic utterance delivered in isolation. These deflections were essentially confined in time to 500 ms after the onset of the acoustic event, which we refer to as AW2. The presence during AW1 of the lower face and its associated movements in the absence of the acoustic utterance, resulted in an average ERP with little or no recognizable voltage deflections. These results are not interpreted to mean that PLST does not respond to a visual stimulus presented alone. The trigger used to synchronize the average waveform was related to acoustic syllable onset and not to the earlier appearance of either the initial visual fixation image or the static lower face. Lip movements, which began 432 ms (nominally) before syllable onset and which are the salient cues to seen speech in this study, are a series of visual events with a gradual onset and progression. Hence this visual stimulus may not evoke the time locked activity necessary for a detectable averaged ERP. Instead, one would expect an ERP associated with lip movement, if present at all, to be progressively elaborated and necessarily small (e.g. Besle et al., 2004).

For any one subject, ERP deflections in response to AV speech ($A_{da}V_{da}$) were similar, though not identical, in appearance to those seen in response to the acoustic syllable presented alone (see Figs. 2, 4 and 6). In addition, the relationship among the three average ERPs evoked by

their corresponding stimulus types (e.g. Fig. 4) clearly differed among subjects. As will be described subsequently, this subjective impression was tested using comparisons between stimulus types and related to the laterality of the hemisphere from which the recordings were obtained.

Proportion of sites exhibiting significant effects

We first compare, for each subject in the study, the percentage of electrode sites on the grid that exhibited significant differences for each of the five contrasts within each of the three AWs. Table 2 shows the form of the response model along with each of the five planned contrasts and the ERP vectors being tested by each of them. Fig. 5 shows in each panel the proportion of significant recording sites for each subject during AW1 (blue), AW2 (red) and AW3 (green) as a function of the five contrasts. Panels A–C represent data from the speech-dominant hemispheres as determined by WADA testing. Panel D shows data from the subject (L122) for which the hemispheric speech dominance was uncertain. Panels E–H represent results obtained from non-speech-dominant hemispheres in the remaining four subjects. As described in Experimental Procedures, subject R127 (G) had a frontal lobe lesion, whereas subject R98 (E) had a previous resection of the inferior and middle temporal gyri.

Several features in these data that are common to all subjects stand out in Fig. 5. First, during AW1 the percentage of significant electrode sites was low ($<4\%$) for all of the five contrasts. This is the time epoch just prior the onset of the audible syllable during which the face was present and lips began to move. This result indicates that there was little, if any, time locked activity generated by the presence alone of the lower face and moving lips, and that the MANOVA approach we used was relatively insensitive to the apparent random fluctuations in the EEG prior to acoustic stimulation. Second, all subjects exhibited significant effects during AW2, although the proportions of significant recording sites were demonstrably greater on speech-dominant hemispheres. The third general feature is that the proportions of significant electrode sites detected in AW3, like AW1, were either negligibly small or zero when compared with their counterparts in AW2, and they showed no clear relationships to the contrasts. Thus, we have restricted subsequent illustrations of results to AW2.

Results of contrast 1 (see Table 2, row C1) show that for speech dominant hemispheres a relatively high proportion of recording sites exhibited a significant effect. Far fewer significant sites were identified on non-dominant hemispheres or on the hemisphere for which dominance was uncertain. We interpret these significant effects to mean that at the effective recording sites the response to the auditory utterance /da/ was influenced by the subject’s ability to view the articulation of that utterance. In terms of the current model (Table 2), the significant effect found for this contrast reflects the combined influence of two response vectors: the visual-alone response (\vec{V}_{da}) and the interaction response (\vec{INT}). We are unable to measure

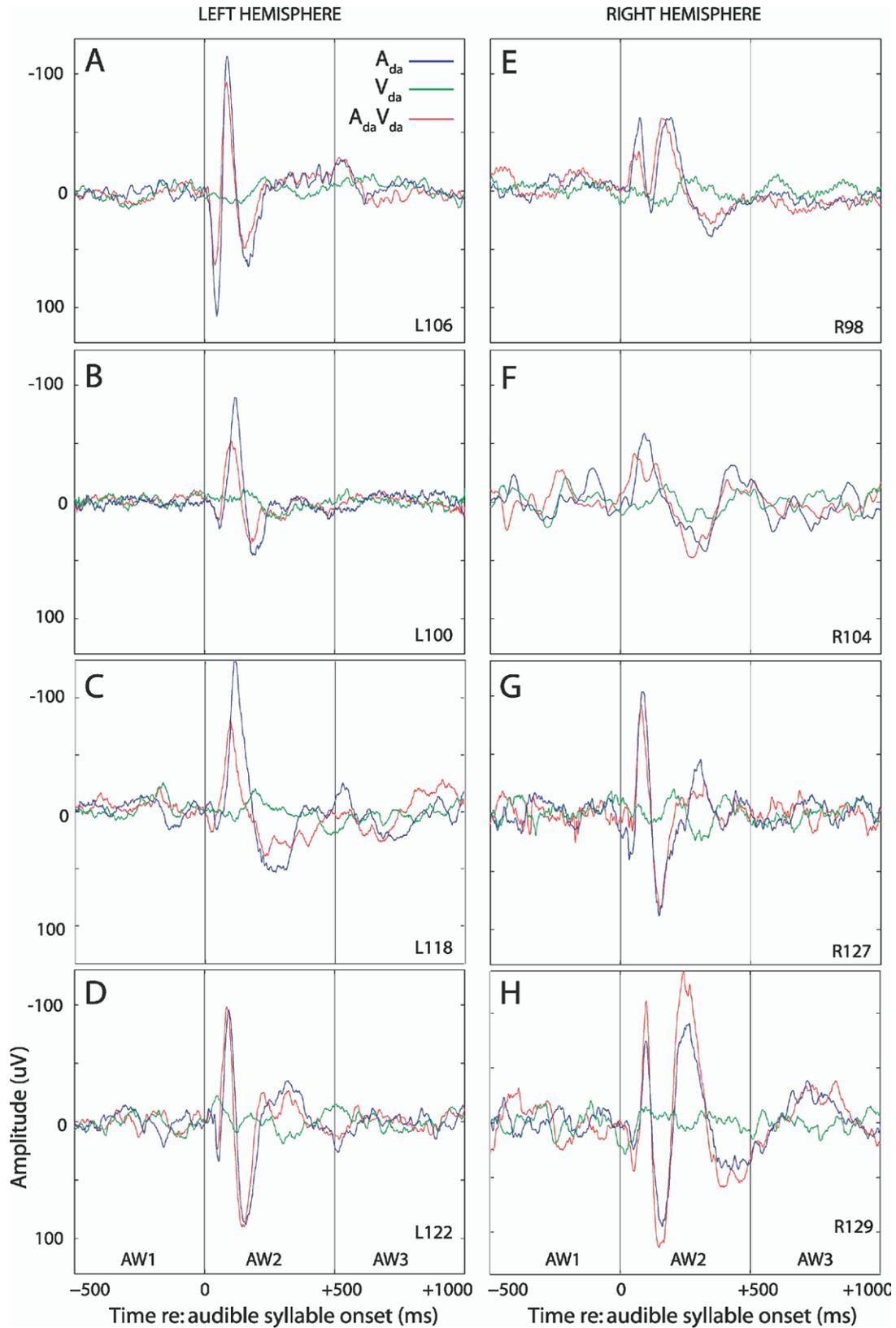


Fig. 4. Average ERPs obtained for three stimulus types. Responses shown from one electrode site (maximal voltage) within area PLST from each subject. Visual alone (V_{da}) stimulus type produces ERPs with minimal voltage deflections from baseline. Auditory alone (A_{da}) and AV speech ($A_{da}V_{da}$) always produce a series of positive and negative deflections beginning shortly (15–50 ms) after audible syllable onset.

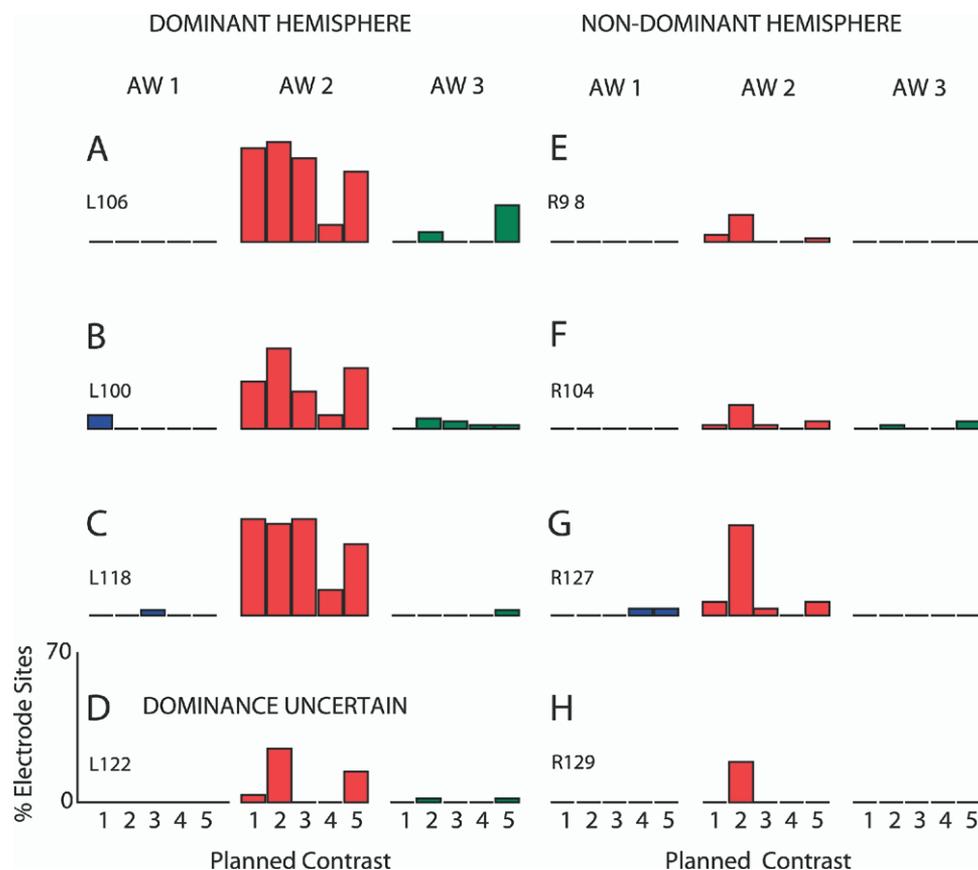


Fig. 5. Percentage of electrode sites with significant effects. Results shown for each subject individually analyzed using MANOVA and multiple-comparison adjustments. The number of electrode sites exhibiting a significant effect in each AW, and for each contrast, was expressed as the proportion of total sites examined.

directly the interaction response vector, but because the visual stimulus presented in isolation evoked little or no apparent time-locked activity during AW2 at most recording sites, a significant effect for C1 may be carried largely by \vec{INT} i.e. the response that is only elicited by a bimodal stimulus. The remaining planned contrasts provide additional information on the contributions of \vec{V}_{da} and \vec{INT} , as well as on the need that there be meaningful or congruent lip movement in order to achieve a significant difference between two stimulus conditions.

The relatively high proportion of sites exhibiting a significant effect for contrast 2 (see Table 2, row C2) reflects the combined influence of the auditory-alone (\vec{A}_{da}) and the interaction responses (\vec{INT}). We assume that for C1 and C2, the interaction response vector term is the same, as \vec{INT} arises from the bimodal AV utterance common to both contrasts. However, the auditory-alone stimulus evoked widespread and robust ERPs bilaterally, while the visual-alone stimulus evoked unremarkable ERPs at the same locations. These observations would suggest that while \vec{INT} was expressed in both contrasts, a significant difference between responses to $\vec{A}_{da}\vec{V}_{da}$ and \vec{V}_{da} stimulus types

in contrast C2 was largely attributable to the auditory-alone response (\vec{A}_{da}).

Contrast 3 (see Table 2, row C3) is similar to C1 in that it tests the difference between responses to AV and V stimulus types. However, in C3 we substituted meaningless (\vec{V}_{gurn}) for congruent (\vec{V}_{da}) lip movements. The results showed a similarly (re. C1) high proportion of significant electrode sites on speech-dominant hemispheres. As with C1, a significant difference is attributable to the combination of a responses \vec{V}_{gurn} and \vec{INT} , with \vec{INT} most likely dominating the effect. The results from contrasts C4 and C5 support this suggestion. First, in C4 a significant effect is attributable to the difference of the two (unremarkable) visual response vectors, \vec{V}_{da} and \vec{V}_{gurn} , and the proportion of significant sites is low. Second, in C5 a significant effect is attributable to the difference of the two auditory responses, \vec{A}_{da} and \vec{A}_{gi} , and the proportion of significant sites is high.

Cortical representation of AV influences

We now turn attention to the question of where on the cortex each significant difference was expressed, knowing that for all subjects the vast majority of effects took place

during the 500 ms window (AW2) after the onset of the acoustic event. Because of the considerable inter-subject variation in gross anatomical landmarks and in location of recording grids with respect to the location of area PLST it was not feasible to pool data and thereby make grand-average comparisons across our subject population without blurring the representational results. We therefore present the individual cortical significance maps for each subject for all contrasts, restricting our description of the distribution of the significant recording sites to AW2 and, with the exception of Fig. 8, to the left hemisphere, as essentially all effects involving the AV-interaction response took place during this time window and within this hemisphere.

Contrast 1: $\overrightarrow{A_{da}V_{da}}$ vs. $\overrightarrow{A_{da}}$

We found consistently that, for any given subject, the response field associated with congruent AV speech ($\overrightarrow{A_{da}V_{da}}$) and that obtained with auditory stimulation alone ($\overrightarrow{A_{da}}$) were characterized by polyphasic ERPs with similar amplitude distributions and time courses. Nevertheless, it is apparent from Fig. 4 that differences between the ERPs evoked with $\overrightarrow{A_{da}V_{da}}$ and $\overrightarrow{A_{da}}$ stimulus types were obtained from some subjects at certain electrode sites.

We begin by presenting in detail data for subject L106 (Fig. 6). In this subject the recording grid was located over the left (speech dominant) hemisphere. The grid covered much of the middle and posterior aspects of the STG, and extended onto the parietal lobe above and the MTG below. AV speech ($\overrightarrow{A_{da}V_{da}}$) as well as auditory-alone ($\overrightarrow{A_{da}}$) stimulation evoked robust responses with the largest Spline–Laplacian ERPs localized to the STG (Fig. 6 B and C). Despite the similarity between response fields, the statistical results indicated that significant differences were represented (Fig. 6A) over essentially the entire response fields (see online Supplementary Data: Results). Thus, it appears that the waveform differences apparent at the single electrode site in Fig. 4A were significant and representative of most other sites in the response fields. Fig. 9 (B and C) presents C1 significance maps from two additional left hemisphere, speech-dominant, cases that are consistent with the results from subject L106.

We interpret the $\overrightarrow{A_{da}V_{da}}$ vs. $\overrightarrow{A_{da}}$ significance map as identifying those cortical locations where the ECoG was influenced significantly by the simultaneous presence of two responses: $\overrightarrow{V_{da}} + \overrightarrow{INT}$. However, the observation that $\overrightarrow{V_{da}}$ response fields were, as a rule, unremarkable in comparison to response fields obtained with a stimulus that included an audible syllable suggests that this statistical effect is largely attributable to \overrightarrow{INT} —the interaction response that is elicited only by a bimodal stimulus.

Only a few recording sites on the non-speech-dominant hemispheres exhibited significant effects for contrast C1 as can be appreciated by the low percentages listed in Fig. 5 (E–H), and hence the statistical maps for these subjects are not shown. Although grid coverage of the STG was less extensive on these right hemispheres, it appears

that the number of the auditory responsive sites detected was sufficient to uncover an AV interaction in these cases, if it were present. Fig. 9D shows the significance map for contrast C1 obtained from the left-hemisphere patient for which speech dominance was uncertain. In this case, the statistical mapping data appear more consistent with the non-speech-dominant, right hemispheres subjects, although L122 also demonstrated areas of cortical dysplasia within the STG.

Contrast 2: $\overrightarrow{A_{da}V_{da}}$ vs. $\overrightarrow{V_{da}}$

In Fig. 7 we present in detail contrast C2 data for the same subject (L106) for which we previously presented C1 results. Whereas AV speech ($\overrightarrow{A_{da}V_{da}}$) evoked robust responses with most easily-recognized Spline–Laplacian ERPs localized to the STG (Fig. 7B), relatively few discernable deflections were seen in the waveforms corresponding to the $\overrightarrow{V_{da}}$ stimulus type that was generated when these same lip movements were presented without the accompanying audible utterance (Fig. 7C). Thus, a Spline–Laplacian ERP to the visual-alone stimulus, if present, was small in amplitude and apparently localized to a few sites at or near where the response to the AV stimulus was most effective. The cortical significance map for contrast C2 (Fig. 7A) could easily have been suggested by comparing by eye the two constituent response fields (i.e. Fig. 7B and C). In this subject, 30 of 64 electrode sites (47%) exhibited a significant difference for contrast C2 with all but a few of them clustered over the STG. This result is remarkably similar in both the proportion (42%) and in the spatial distribution of significant sites obtained with contrast C1 for the same subject.

Significance maps for contrast C2 are shown in Fig. 8 for six additional subjects. Two (A, B) were derived from the left, speech-dominant, hemisphere and one (C) from the left hemisphere where the speech dominance was uncertain. Three maps (D, E, F) were derived from right, non-speech dominant, hemispheres. Like subject L106 (Fig. 7), for all subjects significant electrode sites were commonly observed for this contrast and their locations tended to aggregate over posterolateral STG. In all subjects the recording grid sampled the cortical areas dorsal and ventral to the STG, although significant sites were rarely detected there. The analysis included the two subjects (R127, Fig. 8E: R98) with lesions that altered cortico-cortical inputs from the frontal lobe (R127) or from the cortex lining the ventral bank of the STS (R98).

We interpret the $\overrightarrow{A_{da}V_{da}}$ vs. $\overrightarrow{V_{da}}$ significance map as showing those cortical locations where the ECoG was influenced significantly under the presence of response vectors $\overrightarrow{A_{da}} + \overrightarrow{INT}$ (see online Supplementary Data: Results). These contrast data by themselves, however, do not permit parsing the significance effect between these two responses: $\overrightarrow{A_{da}}$ and \overrightarrow{INT} . Auditory alone ($\overrightarrow{A_{da}}$) stimulation produced a clearly distinguishable response field in this subject and all others in this study. As stated earlier, we assume that the \overrightarrow{INT} response inferred from the results of

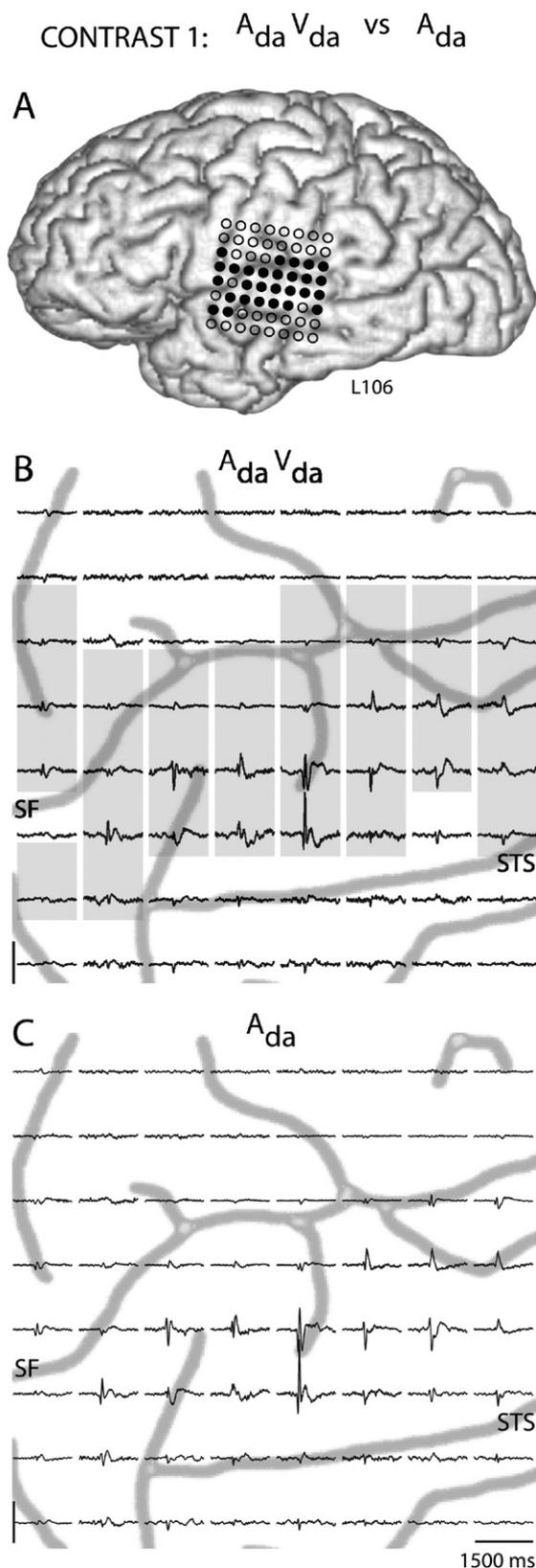


Fig. 6. Comparison of spatial maps of Spline–Laplacians for Contrast 1. (A) Filled circles mark the recording sites at which C1 was significant in AW2. (B) Average Spline–Laplacian waveform at each of the

C1 is the same as that seen with C2, as the AV stimulus was the same for both C1 and C2. If, however, the \vec{A}_{da} response vector was very large as compared with \vec{INT} then these significance maps could be interpreted as arising mainly from the \vec{A}_{da} response. This interpretation would be consistent with the demonstrated C2 cortical significance maps in both left and right hemispheres in Fig. 8 and the observation that response fields for AV speech ($A_{da} V_{da}$) were similar in extent to those obtained with auditory stimulation alone (A_{da}), regardless of hemisphere.

Contrast 3: $A_{da} V_{gurn}$ vs. A_{da}

From Table 1 we see that, according to our model, congruent AV speech ($A_{da} V_{da}$) could elicit as many as three

possible response vectors ($\vec{A}_{da} + \vec{V}_{da} + \vec{INT}$). Similarly, incongruent AV speech ($A_{da} V_{gurn}$), produced by substituting a gurning motion for the natural movement of the speaker's

mouth, also could elicit three vectors ($\vec{A}_{da} + \vec{V}_{gurn} + \vec{INT}$). If, as suggested above, visual-alone stimulation produced a small or negligible response vector, compared with the interaction term, then contrast C3 will test whether congruent lip movement is required for significant effects. Furthermore,

if the interaction response (\vec{INT}) does not differ substantially between the $A_{da} V_{gurn}$ and $A_{da} V_{da}$ stimulus types, then the significance maps for contrasts C3 and C1 should bear a close resemblance to each other. The applicable cortical significance maps are shown in Fig. 9 for the four left hemisphere subjects in our population. Images are arranged to facilitate direct comparison for each subject between the significance maps for contrasts C1 (left column) and C3 (right column). It seems remarkable that these map pairings are nearly identical in the number and location of significant recording sites. This finding provides further support for our contention that the significant effect reliably detected at these electrode sites is carried mainly

by the AV interaction (\vec{INT}) and suggests that the effect is not dependent on congruent lip movement. These suggestions are further supported by results from contrasts C4 and C5.

Contrast 4: $A_{da} V_{da}$ vs. $A_{da} V_{gurn}$

From Table 1 we see that, with our model, congruent AV speech was considered to elicit response components

($\vec{A}_{da} + \vec{V}_{da} + \vec{INT}$), that differed only in the visual term from those vectors resulting from incongruent AV speech ($\vec{A}_{da} + \vec{V}_{gurn} + \vec{INT}$). Thus, contrasting these two stimulus

original recording sites elicited by congruent AV speech ($A_{da} V_{da}$). Gray rectangles replace filled circles. (C) Average Spline–Laplacian waveforms elicited by unimodal Auditory speech (A_{da}). The SF and STS are shown in gray on the spatial maps of waveforms. The ordinate scale (vertical line: -200 to $+200 \mu V/cm^2$) is common to both maps and the abscissa scale includes all three AWs.

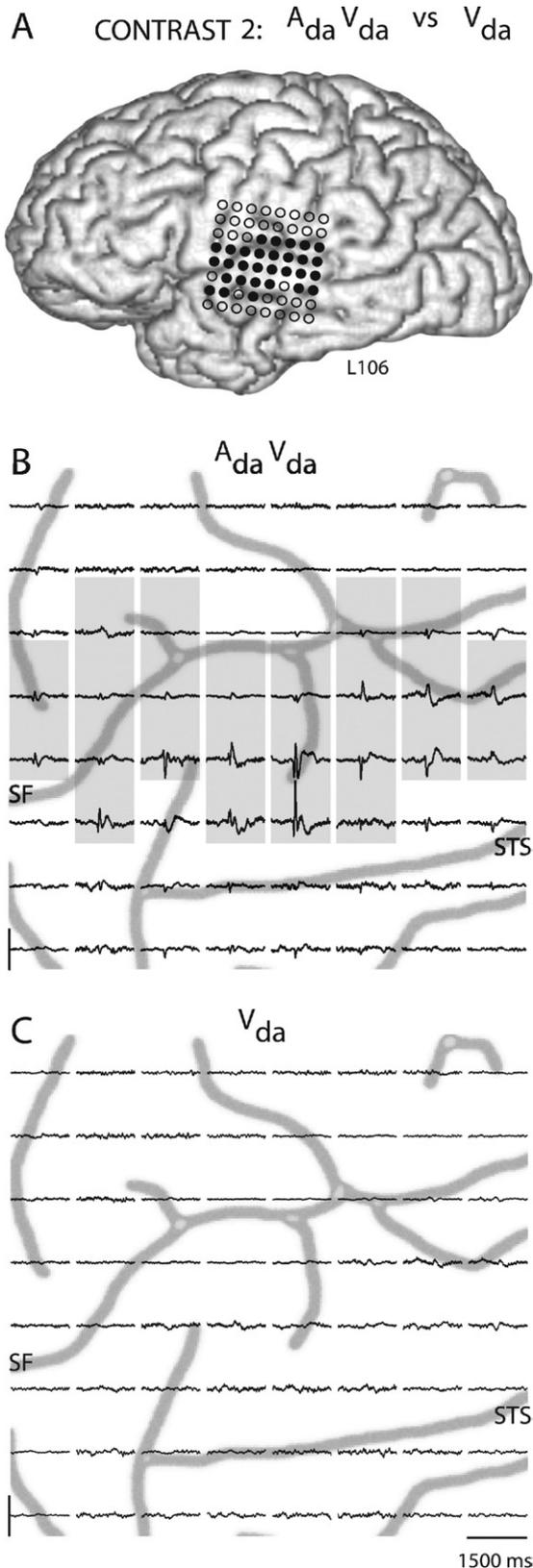


Fig. 7. Comparison of spatial maps of Spline–Laplacians for Contrast 2 in subject L106: (A) Filled circles mark the recording sites at which this contrast was significant in AW2. (B) Average Spline–Laplacian

types (Contrast 4, Table 2) resulted in testing the difference between the two visual response vectors ($\vec{V}_{da} - \vec{V}_{gurn}$). The AV interaction response (\vec{INT}) had cancelled out since the current model makes no distinction between the \vec{INT} vector elicited by $A_{da}V_{da}$ and the interaction response resulting from $A_{da}V_{gurn}$. If, as shown earlier, visual-alone stimulation produced relatively small and scarce response vectors (see Figs. 4 and 7), then testing an effect that depended on their difference was expected to produce a significant outcome at few, if any, electrode sites. This prediction was upheld as shown by the significance maps presented in Fig. 10. On the speech-dominant left-hemispheres (Fig. 10A–C) there were no more than five significant recording sites and on the dominance-undetermined hemisphere (Fig. 10D) and on the non-speech-dominant right hemispheres (not shown) there were none. Those significant sites were within the spatial domain of AV interactions seen with contrasts C1 and C3, which suggests that differences in the visual-alone response vectors, though small, could still produce measurable effects.

Contrast 5: $A_{da}V_{da}$ vs. $A_{gi}V_{da}$

Incongruent AV speech ($A_{gi}V_{da}$), produced by substituting the audible syllable/gi/ for the syllable /da/ while retaining the natural movements of the speaker's mouth for syllable /da/, elicits three response vectors, $\vec{A}_{gi} + \vec{V}_{da} + \vec{INT}$ (Table 1). These response components differ only in the auditory term from those evoked by congruent AV speech, $\vec{A}_{da} + \vec{V}_{da} + \vec{INT}$; assuming again that the interaction vectors are identical. Therefore, contrasting these two stimulus types (contrast C5, Table 2) resulted in testing the difference between the two auditory responses ($\vec{A}_{da} - \vec{A}_{gi}$), as the AV interaction canceled out. Unlike visual-alone stimulation, auditory-alone stimulation with any CV syllable was typically capable of evoking ERPs in area PLST similar to those depicted in Fig. 3B. Therefore, a significant effect for contrast C5 would depend on neural processing that discriminated between acoustic signals for da/ and gi/. Fig. 11 illustrates the cortical significance maps corresponding to contrast C5 for the four left-hemisphere subjects in our population. Clearly, in speech-dominant hemispheres (Fig. 11A, B, C) the number of electrode sites at which the contrast was significant suggests that the acoustic difference in the syllables was discriminated. The significance map obtained from the left-hemisphere patient for which speech dominance was uncertain (D) does not show this effect. Similarly, the statistical results from the four non-speech

waveform at each of the original recording sites elicited by congruent AV speech ($A_{da}V_{da}$). Gray rectangles replace filled circles. (C) Average Spline–Laplacian waveforms elicited by unimodal Visual speech (V_{da}). The locations of the SF and STS are shown in gray on the spatial maps of waveforms. The ordinate scale is common to both maps (vertical line: -200 to $+200 \mu V/cm^2$) and the abscissa scale includes all three AWs.

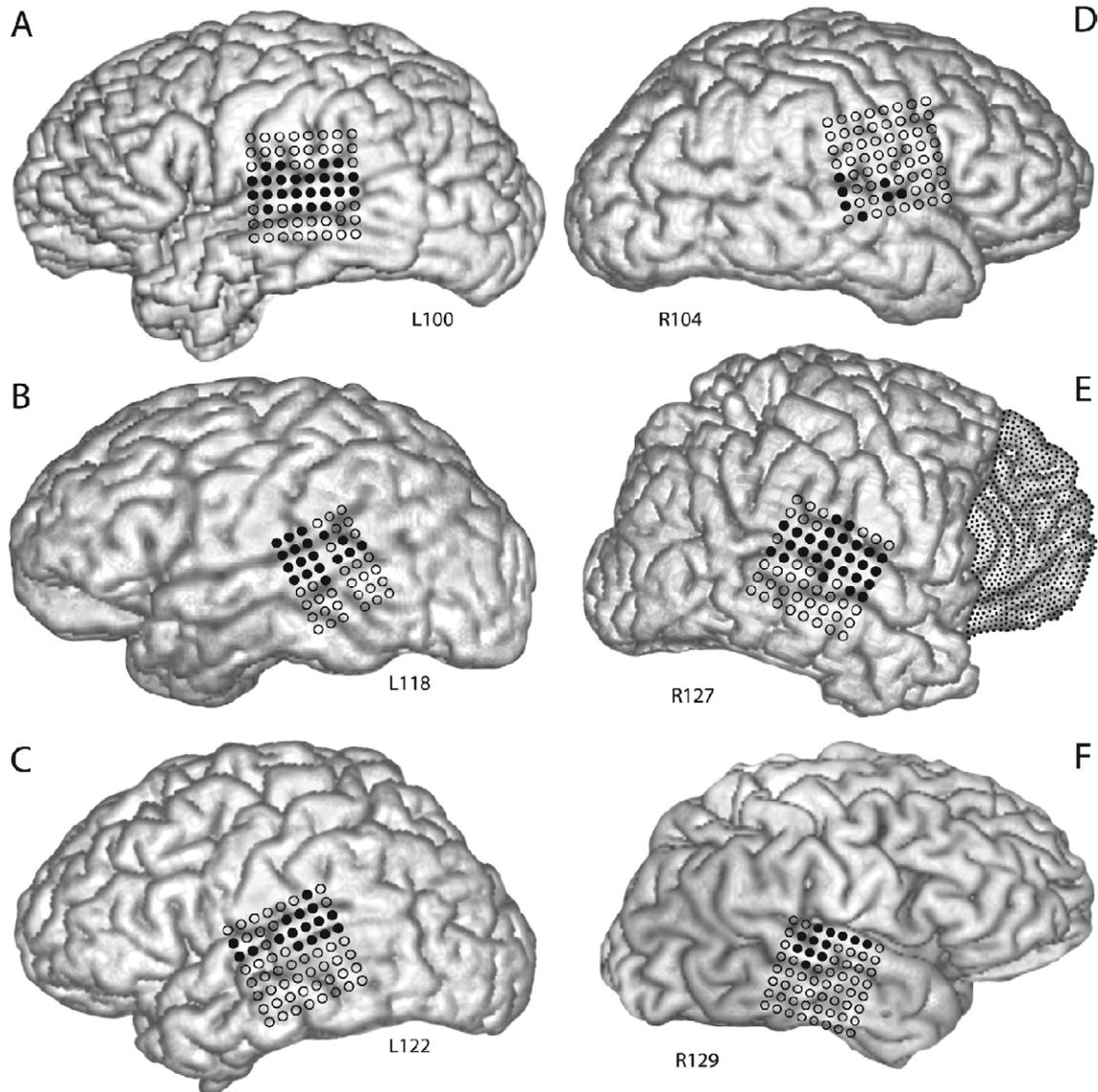
CONTRAST 2: $A_{da} V_{da}$ vs V_{da} 

Fig. 8. Cortical significance maps from three left (A–C) and three right (D–F) hemisphere subjects for contrast C2. Filled circles mark sites where this contrast was significant in AW2. Stippled area in (E) shows the extent of frontal lobe damage sustained earlier by subject R127 (see text for further description).

dominant, right hemisphere subjects indicate proportions of significant effects much smaller (Fig. 5E–H) than those observed in speech-dominant hemispheres (Fig. 5A–C).

For every subject in this study, the proportion of significant electrode sites detected for contrast C5 was approximately equal to the proportion obtained with either contrast C1 or C3 (see Fig. 5). For contrasts C1 and C3, we argued that the significant effect could be largely attributed to an AV interaction response. Furthermore, the cortical significance maps for contrast C5 in speech-dominant hemispheres (Fig. 11A, B, C) are nearly co-extensive with the significance maps for these subjects

under either contrast C1 (Fig. 9A, B, C) or C3 (Fig. 9E, F, G). Taken together these findings suggest that overlapping territories in posterolateral STG on the speech-dominant hemisphere are engaged in the neural processing associated with both AV interactions and acoustic syllable discrimination.

Test for laterality effect

Since we did not have an a priori hypothesis to test for a laterality effect, we took an exploratory approach and performed nonparametric tests for location differences across a one-way classification. Our data were classified into two groups: those subjects who had electrodes placed on the

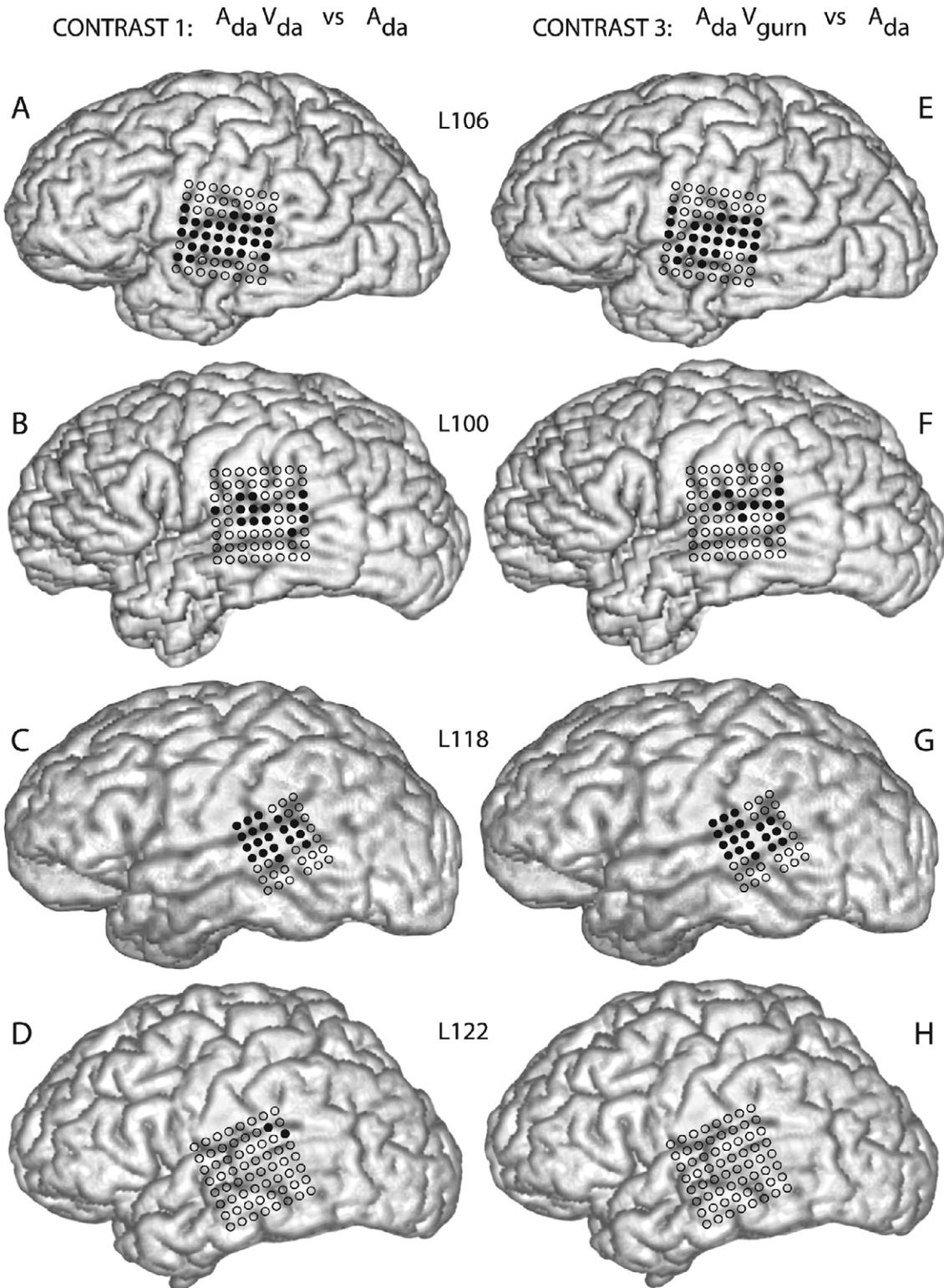


Fig. 9. Cortical significance maps for contrasts C1 (A–D) and C3 (E–H) for the four left hemisphere subjects. Filled circles mark sites where the contrast was significant in AW2. The left hemisphere of three subjects (L106, L100, L118) was speech dominant. Speech dominance of L122 was not known. Maps for C1 and C3 for the four right hemisphere subjects exhibited zero to four significant sites (not shown).

temporal lobe of the LEFT hemisphere and those who had electrodes in the RIGHT hemisphere. The response vari-

able was the F -value calculated from Wilkes' lambda that resulted from planned contrasts C1 and C3 since a signif-

CONTRAST 4: $A_{da} V_{da}$ vs $A_{da} V_{gurn}$

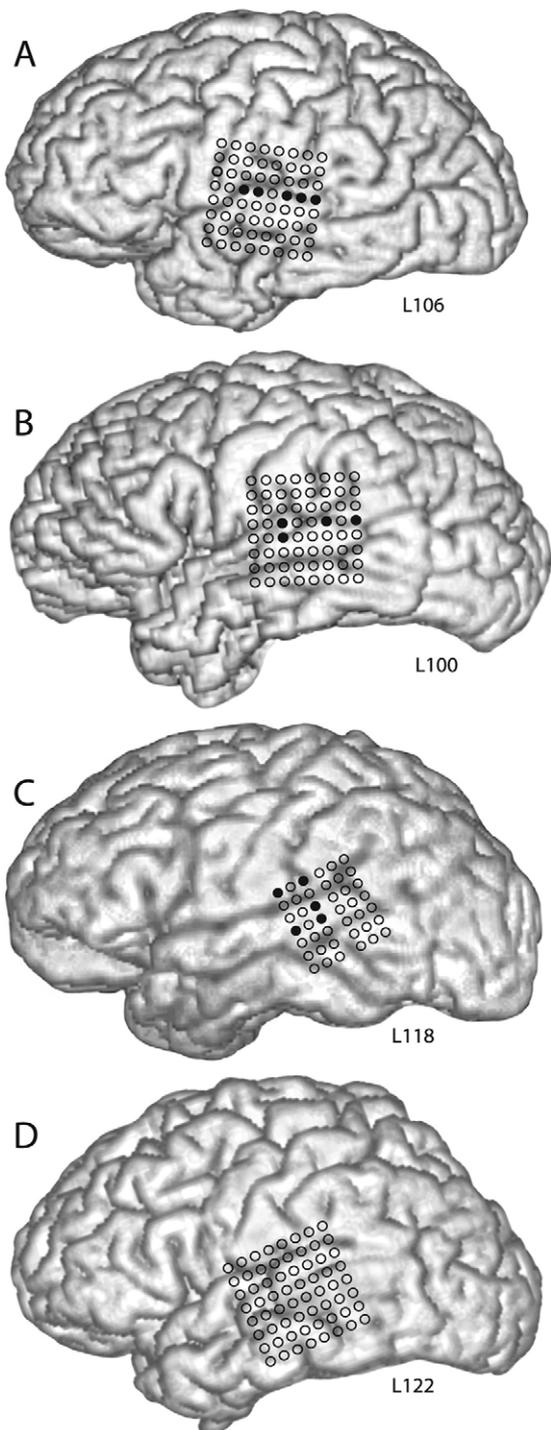


Fig. 10. Cortical significance maps for contrast C4 for the four left hemisphere subjects. Meaningless mouth movement (gurning) substituted for the natural movement of the speaker's mouth in articulating the syllable /da/. Filled circles mark sites where this contrast was significant in AW2. No significant sites were found in the four right hemisphere cases.

icant effect in these contrasts was attributed to an AV interaction that was relatively larger than either visual-

alone response and therefore dominated the effect. Our goal was to decide whether the location and distribution of this response variable differed for the two populations. The nonparametric procedure (Procedure NPAR1WAY, SAS 9.1) performed a test for location using the Van der Waerden (VW) scores in a simple linear rank statistic. VW scores are the quantiles of a standard normal distribution. These scores are also known as *quantile normal scores* and are powerful for normal distributions. The VW two-sample test statistic was significant ($Z=8.34$, $P<0.0001$) indicating rejection of the null hypothesis of no difference between LEFT and RIGHT hemisphere populations. A similar effect was found using the raw F -values as the input scores ($Z=7.96$, $P<0.0001$). To test the hypothesis that the two groups of observations have identical distributions, the procedure provides empirical distribution function statistics, including the two-sample Kolmogorov-Smirnov test statistic. The result for the KS two-sample test was significant ($KS=3.31$, $P<0.0001$), which indicates rejection of the null hypothesis that the F -value distributions were identical for the two levels of hemisphere. The same pattern of statistically significant effects were observed when using exact statistics based on Monte Carlo estimation.

DISCUSSION

Speech communication often integrates hearing and seeing, and thus it should not be surprising to find that human hemodynamic (Callan et al., 2003, 2004; Calvert et al., 1997, 1999, 2000; Calvert, 2001; MacSweeney et al., 2002; Pekola et al., 2005; van Atteveldt et al., 2004), MEG (Mottonen et al., 2002; Sams et al., 1991), and EEG (Callan et al., 2001; Besle et al., 2004; Giard and Peronnet, 1999; Klucharev et al., 2003; Molholm et al., 2002; Murray et al., 2005; van Wassenhove et al., 2005) studies have implicated widespread involvement of superior temporal auditory cortex in AV interactions. Human auditory cortex is made up of multiple fields, however, and with the exception of core area A1 there is still no full agreement on their number and spatial arrangement (Hackett et al., 2001, 2003; Hackett, 2003; Formisano et al., 2003; Sweet et al., 2005; Wessinger et al., 2001). Thus, while the converging hemodynamic, MEG and EEG evidence from human studies points to the STG as playing a role in AV interactions, the question of which of the multiple auditory fields are so involved has not been fully answered by those experiments.

Our major finding here is that AV interactions are represented within physiologically-defined auditory area PLST on the human posterolateral STG. The AV representation overlaps extensively with PLST on the speech-dominant hemisphere, but is hardly in evidence on the non-dominant hemisphere. The results were derived from ERPs to heard and seen speech recorded directly from lateral temporal cortex using pial-surface electrodes implanted in neurosurgical patients undergoing diagnosis and treatment of medically intractable epilepsy. The ERP maps obtained were first processed by the Spline-Laplacian technique, which improved the spatial resolution of local cortical sources and de-emphasized deep brain sources and/or coherent

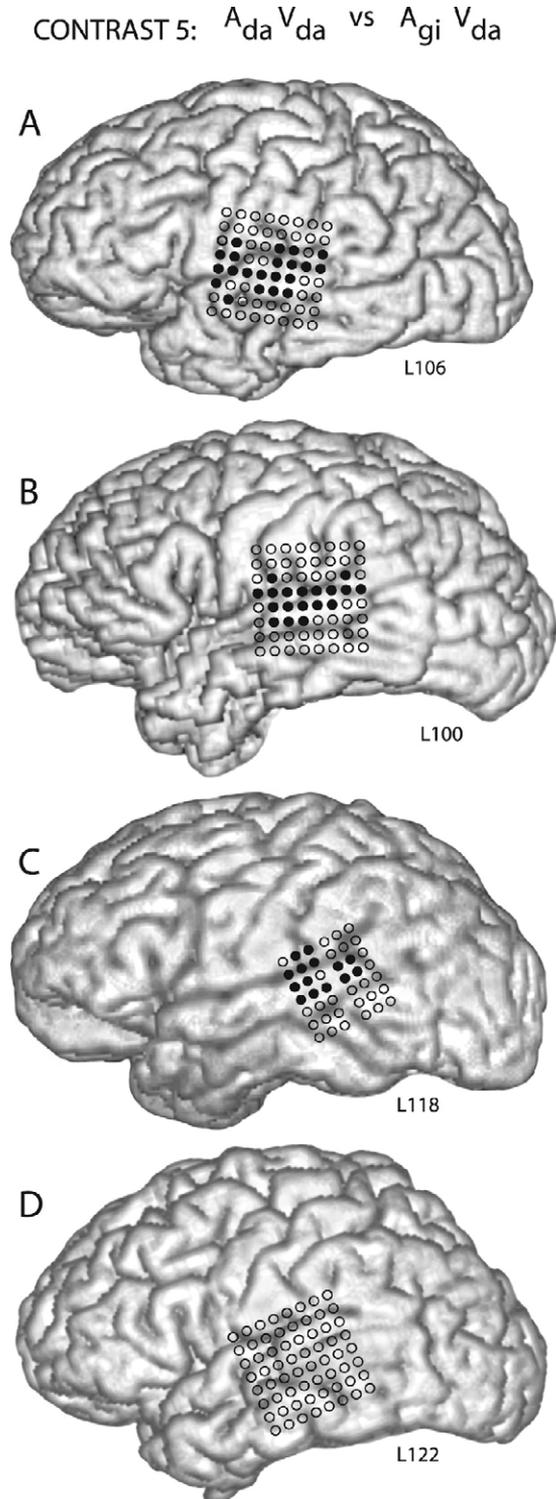


Fig. 11. Cortical significance maps for contrast C5 for the four left hemisphere subjects. The audible syllable /gi/ substituted for syllable /da/ while retaining the natural movements of the speaker's mouth for syllable /da/. Filled circles mark sites where this contrast was significant in AW2. Few significant sites were found in the four right hemisphere cases.

sources distributed over large cortical areas. Data were then subjected to a MANOVA analysis (see online Supple-

mentary Data: Discussion) to test whether the Spline–Laplacian ERPs differed among three experimental factors, including the stimulus type, electrode site and AW.

In our cortical significance maps, AV interactions were most frequently localized to electrode sites in area PLST, although a few sites were routinely located above the SF in parietal cortex. These parietal sites are typically located at the physical edge of the recording grid and, therefore, may represent a computational artifact (see Experimental Procedures: Spline–Laplacian transformation). Cortex of the MTG and posterior STS has been identified in human fMRI studies as sites of AV integration to both non-speech (Beauchamp et al., 2004) as well as speech tokens (Callan et al., 2004). The latter authors proposed that the multi-sensory effects noted in their study might have their genesis in the biological motion that occurs during natural speech articulation. In this regard, the posterior STS region, but apparently not adjacent STG, is reported to exhibit robust hemodynamic activation in human subjects viewing eye and mouth movements without an accompanying auditory component (Puce et al., 1998). It seems quite unlikely, however, that the AV interaction we found represented within field PLST arose from activity spreading from multimodal cortex of the STS or from the MTG where in several subjects the recording sites were located. Our spatial filtering likely reduced or eliminated far field influences that may have arisen from possible AV activity in these distant regions. We interpret the fact that we found no evidence of AV interactions from recorded sites on MTG to mean that our coverage of this gyrus was not extensive enough to identify active sites, or that the stimuli used were not of the kind to elicit such an interaction.

AV interactions

Significant differences that arose from using five contrasts between bimodal and unimodal responses were, depending on the contrast, interpreted to reflect differences in the unimodal response, an AV interaction, or some combination of the two. In other words, an AV interaction was identified under a simple model of vector response summation. In testing for significance, the analyses took into account the entire waveform within each AW and at each electrode site. Because the aim of the study was to test the hypothesis that AV interactions occurred within a known auditory field, this approach was chosen to identify *where* on the posterolateral temporal cortex significant differences were expressed during each of three AWs. The question of *when* the interactions occurred is left for follow-up studies and analyses.

The major finding of an AV interaction came from the results of contrast C1 ($\vec{A}_{da} V_{da}$ vs. \vec{A}_{da}). Although the auditory-alone stimulus and the congruent-AV stimulus both evoked ERPs that were similar in waveform and in spatial distribution on the STG, our analysis often revealed significant differences at many of the same sites that were responsive to auditory stimulation in the speech-dominant hemisphere. According to our model, these differences could have arisen from a visual response, from an AV

interaction response or from a combination of the two. The visual stimulus when presented in isolation evoked little obvious time-locked response during any AW. If an ERP was associated with the visual-alone stimulus during AW2, it appeared to be very low in amplitude and confined to but one or a few sites within the significance map. From this we inferred that the visual stimulus contributed little to the results of contrast C1, and that it was an interaction response vector that was mainly responsible for any significant effects. Our current model incorporated an AV interaction response that did not depend upon the particular form of the bimodal stimulus (e.g. $A_{da}V_{da}$ or $A_{da}V_{gurn}$ or $A_{gi}V_{da}$) that evoked the interaction. This formulation was supported by the remarkable similarity between significance maps for contrasts C1 ($\vec{A}_{da}V_{da}$ vs. \vec{A}_{da}) and C3 ($\vec{A}_{da}V_{gurn}$ vs. \vec{A}_{da}) since it was argued that the effect for both contrasts was dominated by this same interaction vector. The finding of an AV interaction under both speech conditions is in accord with results from a recent scalp ERP study in which the use of AV speech stimuli provided evidence for multisensory interactions regardless of whether the speech tokens were congruent or incongruent (van Wassenhove et al., 2005).

When we apply our model to contrast C4, however, which tested the difference between response vectors \vec{V}_{da} and \vec{V}_{gurn} , we conclude that not only were there two visual response vectors at a few electrode sites but that, however small, the difference between these vectors was significant. This conclusion follows from the model's formulation of the AV interaction associated with the congruent ($A_{da}V_{da}$) stimulus type as identical to that associated with the incongruent ($A_{da}V_{gurn}$) type. Nevertheless, regardless of the support for this formulation discussed above, we cannot rule out the possibility that the interaction vector may differ depending on the composition of the AV stimulus. Such differences might reasonably be expected to occur for those AV stimulus types that are known to produce robust auditory illusions (e.g. the McGurk effect) or deviance potentials, and the model can be easily re-configured to reflect this dependence. Limitations of recording time prevented us from employing additional stimulus types to address more fully whether and to what extent the AV interactions corresponding to congruent and incongruent cases might, indeed, be unique.

It is not too surprising to find that visual-alone stimuli failed to elicit ERPs more comparable to those evoked by a stimulus containing a causal auditory event. In our current experimental design ERPs were synchronized to effective sounds having relatively rapid onset times. Lip movements, on the other hand, had gradual onsets and progression times and, therefore, were not optimal stimuli for eliciting precise time-locked activity typically revealed by the average ERP. Alternative approaches that do not depend on precise trial-by-trial synchronization (e.g. spectral analysis) may prove more effective in revealing cortical responses to lip movement.

The fact that in humans robust ERPs to AV stimuli are recorded on posterolateral STG indicates that this cortex

receives highly synchronized stimulus-evoked afferent volleys. It is certainly conceivable that the AV interactions we observed in ERPs from posterolateral STG in human are the result of AV information arriving over cortico-cortical pathways. In Old-World monkeys, auditory parabelt on the posterior STG has connections not only with other auditory fields of the temporal lobe, but with regions in the prefrontal cortex (Hackett et al., 1998b, 1999; Molinari et al., 1995; Romanski et al., 1999a,b) and cortex forming the banks of the STS (Ward et al., 1946; Seltzer and Pandya, 1978, 1994; Galaburda and Pandya, 1983; Petrides and Pandya, 1988; Hackett et al., 1998a) considered to have multisensory function. Polysensory STS receives its input from visual cortex and from polysensory thalamic nuclei (Seltzer and Pandya, 1978, 1994; Desimone and Ungerleider, 1986). Additionally, afferent activity arising in visual cortex may be transmitted to the posterior STG directly (Foxe and Schroeder, 2005; Nascimento-Silva et al., 2005; Schroeder et al., 2003). Both the highly synchronized afferent volley evoked by an AV stimulus and the relatively short latency of the onset of the resulting ERP suggest, however, that direct sources of the earliest PLST input at least are the multimodal areas of the thalamus. In the non-human primate, auditory belt and parabelt cortex, unlike the auditory core, receive their thalamo-cortical input mainly from the thalamic nuclei that lie outside of the lateral lemniscal route. These nuclei include the dorsal and magnocellular divisions of the medial geniculate body, supragenicular nucleus, nucleus limitans and medial and central pulvinar (Rauschecker et al., 1997; Hackett et al., 1998b), all of which are well represented in the human thalamus (Winer, 1984). Non-lemniscal thalamus receives multisensory convergent input from visual, somatic sensory and auditory areas of the brainstem and midbrain (Benedek et al., 1997; Graybiel, 1972; Jones, 1985; LeDoux et al., 1987; Wepsic, 1966; Winer, 1992). Hence the most parsimonious interpretation of our results would be that AV interactions recorded in human posterolateral STG are the result of converging cortico-cortical and thalamo-cortical unisensory and/or polysensory inputs.

The non-speech dominant hemispheres of two subjects had sustained lesions of multisensory cortical areas that have known connections with the posterolateral STG (Barbas, 1988; Hackett et al., 1998b, 1999; Molinari et al., 1995; Petrides and Pandya, 1988; Romanski et al., 1999a,b; Seltzer and Pandya, 1978, 1994; Ward et al., 1946). One subject (R98) had undergone resection of the inferior and middle temporal gyri and anterior pole of the STG in an earlier epilepsy procedure. The second (R127) had sustained a traumatic head injury that severely damaged his right frontal lobe. In both cases, responses in PLST to any stimulus type containing an audible syllable provided robust ERPs that appeared similar to those evoked from any other right hemisphere subject. Furthermore, cortical significance maps resulting from all tested contrasts differed in no substantial way between lesioned and non-lesioned cases. Apparently any significant effects that we observed in lesioned cases did not require the presence of either intact

cortico-cortical afferent or efferent pathways connecting these multisensory cortical fields.

In human studies that employed natural AV speech stimuli, significant AV effects can be strongly lateralized to the left hemisphere (Callan et al., 2004; Calvert et al., 2000; Calvert, 2001; MacSweeney et al., 2002; Pekkola et al., 2005). Speech perception is often held to be specialized in the dominant (typically left) hemisphere (Liberman et al., 1967), as speech perception is closely linked to highly lateralized language processing (Geschwind, 1970; Binder et al., 1997). Evidence has accumulated over the past several years, however, that the numerous processes involved in temporal and spectral analyses of the speech signal are mediated bilaterally (Hickok and Poeppel, 2004; Norris and Wise, 2000) although integration time constants may differ between hemispheres (Poeppel and Hickok, 2004; Zatorre et al., 2004). We have not been able to test this latter suggestion directly because of clinical considerations, but it is the case that PLST on either the right or left hemisphere is activated robustly by a wide range of acoustic signals, including CV syllables (see also Binder et al., 2000). Although bilateral activation of PLST holds for acoustic signals presented alone, it appears from our results that the AV interaction occurring when an audible syllable is paired with the visual image of the moving mouth that uttered that sound is disproportionately represented on the speech-dominant hemisphere. Indeed, we saw very little evidence of AV interaction in the non-dominant hemisphere. Our conclusion on this is still tentative as our sample size is quite small, and we are unable to study both the left and right hemispheres in the same subject. Nonetheless, this finding is consistent with those from fMRI studies which identified STS cortex only in the speech dominant left hemisphere as the site with the highest specificity for bimodal integration (Callan et al., 2004; Calvert et al., 2000; Calvert and Lewis, 2004).

PLST—posterior lateral STG

Area PLST is physiologically differentiated from, but functionally connected to, auditory cortex on the supratemporal plane (Brugge et al., 2003, 2005; Howard et al., 2000). From the standpoint of major fissural landmarks, PLST appears to correspond to, or overlap with, the posterior portion of cytoarchitectonic areas 22 of Brodmann (1909), area TA of von Economo (1929), areas PaAe, PaAi and Tpt of Galaburda and Sanides (1980), and chemoarchitectonic area STA of Rivier and Clarke (1997) and Wallace et al. (2002). Where (or whether) PLST stands in a postulated three-tier hierarchical model, which is based largely on anatomical and electrophysiological findings in monkey (Kaas and Hackett, 1998; Rauschecker and Tian, 2000) but for which there is also evidence from functional imaging studies in human (Formisano et al., 2003; Wessinger et al., 2001) is another question.

The lateral surface of the STG is associated with auditory language processing with the posterior extent traditionally referred to as Wernicke's area. Structural lesions or electrical disruption of this area in humans is associated with impairments in phonologic processing, auditory com-

prehension, word/sentence repetition, and simpler forms of acoustic–phonetic processing (Boatman et al., 1995; Crone et al., 2001; Karbe et al., 1998; Lesser et al., 1986; Naeser et al., 1990; Selnes et al., 1985). Thus, these functional observations provide evidence for both belt and parabelt proposed organizational features. In this context, Boatman (2004) has suggested that the organization of lateral STG is not compatible with the traditional single anterior–posterior division scheme, but rather with a structure composed of multiple, functional subdivisions that together support speech perception. We propose that PLST is one such subdivision.

We have tentatively considered PLST a homolog of all or part of the auditory parabelt of monkey (Hackett et al., 1998a). Equally plausible is the possibility that PLST represents a belt field perhaps homologous to the middle lateral area in monkey (Rauschecker et al., 1995; Rauschecker and Tian, 2004). However, more recent cyto- and chemo-architectonic findings in human report that the belt/parabelt does not extend onto the posterior lateral surface of the STG, a region said to be occupied by area Tpt (Galaburda and Sanides, 1980; Sweet et al., 2005). Sweet and colleagues (2005) have argued forcefully that the parabelt in monkey corresponds to only a portion of the planum temporale in humans.

Studies on non-human primates, however, shed light on areas of multimodal representation in human only to the extent that the organization of auditory cortex in human is known and the homologies between human and monkey auditory cortical fields are understood. Our study may help to reconcile these proposed homologies by providing direct electrophysiological recordings from non-core auditory cortical fields in humans that can be compared with data from monkey. As Hackett (2003) points out, homologies beyond the auditory core remain tenuous, especially for parabelt regions, as this cortex may have undergone expansion in humans and thus may include additional areas not accounted for at present. It has been posited that speech processing in the human may involve a ventral serial processing stream that arises bilaterally in the core auditory fields and proceeds to engage multiple belt and parabelt fields located more laterally on the supratemporal plane and STG (Hickok and Poeppel, 2004; Rauschecker and Tian, 2000). Systematic electrical stimulation mapping studies have clearly identified posterolateral STG as an area containing circuitry that critically supports speech perception (reviewed by Boatman, 2004). Our observation of AV interactions being represented in PLST implicates this field as playing a role in processing heard and seen speech at an early stage of speech processing.

Acknowledgments—This work was supported by National Institutes of Health Grants DC-04290, DC-00657, DC-00116, HD-03352, and M01-RR-59 (General Clinical Research Centers Program) and by the Hoover Fund and Carver Trust.

REFERENCES

Arezzo JC, Vaughan HG Jr, Kraut MA, Steinschneider M, Legatt AD (1986) Intracranial generators of event related potentials in the

- monkey. In: *Evoked potential frontiers of clinical neuroscience*, Vol. 3 (Cracco RQ, Bodis-Wollner I, eds), pp 174–189. New York: Alan R. Liss.
- Barbas H (1988) Anatomic organization of basoventral and mediodorsal visual recipient prefrontal regions in the rhesus monkey. *J Comp Neurol* 276:313–342.
- Barracough NE, Xiao D, Baker CI, Oram MW, Perrett DI (2005) Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *J Cogn Neurosci* 17:377–391.
- Baylis GC, Rolls ET, Leonard CM (1987) Functional subdivisions of the temporal lobe neocortex. *J Neurosci* 7:330–342.
- Beauchamp MS, Lee KE, Argall BD, Martin A (2004) Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* 41:809–823.
- Benedek G, Perény J, Kovács G, Fischer-Szátmári L, Katoh YY (1997) Visual, somatosensory, auditory and nociceptive modality properties in the feline supragenulate nucleus. *Neuroscience* 78:179–189.
- Benevento LA, Fallon J, Davis BJ, Rezak M (1977) Auditory-visual interaction in single cells in the cortex of the superior temporal sulcus and the orbital frontal cortex of the macaque monkey. *Exp Neurol* 57:849–872.
- Besle J, Fort A, Delpuech C, Giard MH (2004) Bimodal speech: early suppressive visual effects in human auditory cortex. *Eur J Neurosci* 20:2225–2234.
- Binder JR, Frost JA, Hammeke TA, Cox RW, Rao SM, Prieto T (1997) Human brain language areas identified by functional magnetic resonance imaging. *J Neurosci* 17:353–362.
- Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET (2000) Human temporal lobe activation by speech and nonspeech sounds. *Cereb Cortex* 10:512–528.
- Boatman D, Lesser RP, Gordon B (1995) Auditory speech processing in the left temporal lobe: an electrical interference study. *Brain Lang* 51:269–290.
- Boatman D (2004) Cortical bases of speech perception: evidence from functional lesion studies. *Cognition* 92:47–65.
- Brodman K (1909) *Vergleichende Lokalisationslehre der Grosshirnrinde*. Leipzig, Germany: J. A. Barth.
- Bruce C, Desimone R, Gross CG (1981) Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J Neurophysiol* 46:369–384.
- Brugge JF, Volkov IO, Garell PC, Reale RA, Howard MA (2003) Functional connections between auditory cortex on Heschl's gyrus and on the lateral superior temporal gyrus in humans. *J Neurophysiol* 90:3750–3763.
- Brugge JF, Volkov IO, Reale RA, Garell PC, Kawasaki H, Oya H, Li Q, Howard MA (2005) The posterolateral superior temporal auditory field in humans: Functional organization and connectivity. In: *Auditory cortex: a synthesis of human and animal research* (Konig R et al., eds), pp 145–162. Mahwah, NJ: Erlbaum.
- Callan DE, Callan AM, Kroos C, Vatikiotis-Bateson E (2001) Multimodal contribution to speech perception revealed by independent component analysis: a single-sweep EEG case study. *Brain Res Cogn Brain Res* 10:349–353.
- Callan DE, Jones JA, Munhall K, Callan AM, Kroos C, Vatikiotis-Bateson E (2003) Neural processes underlying perceptual enhancement by visual speech gestures. *Neuroreport* 14:2213–2218.
- Callan DE, Jones JA, Munhall K, Kroos C, Callan AM, Vatikiotis-Bateson E (2004) Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. *J Cogn Neurosci* 16:805–816.
- Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire PK, Woodruff PW, Iversen SD, David AS (1997) Activation of auditory cortex during silent lipreading. *Science* 276:593–596.
- Calvert GA, Brammer MJ, Bullmore ET, Campbell R, Iversen SD, David AS (1999) Response amplification in sensory-specific cortices during crossmodal binding. *Neuroreport* 10:2619–2623.
- Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol* 10:649–657.
- Calvert GA (2001) Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cereb Cortex* 11:1110–1123.
- Calvert GA, Lewis JW (2004) Hemodynamic studies of audiovisual interactions. In: *The handbook of multisensory processing* (Calvert GA et al., eds), pp 483–502. Cambridge: MIT Press.
- Calvert GA, Thesen T, Hansen PC, Singh K, Campbell R, Holliday IE (2005) Changes in cortical oscillatory power during audio-visual speech perception revealed by MEG. *Annual Meeting of the Society for Cognitive Neuroscience* 216 (Suppl S).
- Campbell R, Dodd B (1980) Hearing by eye. *Q J Exp Psychol* 32:85–99.
- Campbell R, Landis T, Regard M (1986) Face recognition and lip-reading: a neurological dissociation. *Brain* 109:509–521.
- Campbell R, MacSweeney M, Surguladze S, Calvert G, McGuire P, Suckling J, Brammer MJ, David AS (2001) Cortical substrates for the perception of face actions: an fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning). *Brain Res Cogn Brain Res* 12:233–243.
- Creutzfeldt OD, Houchin J (1984) Neuronal basis of EEG-waves. In: *Handbook of electroencephalography and clinical neurophysiology*, Vol. 2 (Remond A, ed), pp 2/C/3–2/C/55. Amsterdam: Elsevier.
- Creutzfeldt O, Ojemann G, Lettich E (1989) Neuronal activity in the human lateral temporal lobe. I. Responses to speech. *Exp Brain Res* 77:451–475.
- Crone NE, Boatman D, Gordon B, Hao L (2001) Induced electrocorticographic gamma activity during auditory perception. *Brazier Award-winning article, 2001. Clin Neurophysiol* 112:565–582.
- Desimone R, Gross CG (1979) Visual areas in the temporal cortex of the macaque. *Brain Res* 178:363–380.
- Desimone R, Ungerleider LG (1986) Multiple visual areas in the caudal superior temporal sulcus of the macaque. *J Comp Neurol* 248:164–189.
- Dillon WR, Goldstein M (1984) *Multivariate analysis. Methods and applications*. New York: Wiley and Sons.
- Dodd B (1977) The role of vision in the perception of speech. *Perception* 6:31–40.
- Donchin E (1966) A multivariate approach to the analysis of average evoked potentials. *IEEE Trans Biomed Eng* 13:131–139.
- Formisano E, Kim DS, Di Salle F, van de Moortele PF, Ugurbil K, Goebel R (2003) Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron* 40:859–869.
- Fowler CA (2004) Speech as a supramodal or amodal phenomenon. In: *The handbook of multisensory processing* (Calvert GA et al., eds), pp 189–202. Cambridge: MIT Press.
- Foxe JJ, Morocz IA, Murray MM, Higgins BA, Javitt DC, Schroeder CE (2000) Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Brain Res Cogn Brain Res* 10:77–83.
- Foxe JJ, Wylie GR, Martinez A, Schroeder CE, Javitt DC, Guilfoyle D, Ritter W, Murray MM (2002) Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. *J Neurophysiol* 88:540–543.
- Foxe JJ, Schroeder CE (2005) The case for feedforward multisensory convergence during early cortical processing. *Neuroreport* 16:419–423.
- Fu KM, Johnston TA, Shah AS, Arnold L, Smiley J, Hackett TA, Garraghy PE, Schroeder CE (2003) Auditory cortical neurons respond to somatosensory stimulation. *J Neurosci* 23:7510–7515.
- Galaburda AM, Sanides F (1980) Cytoarchitectonic organization of the human auditory cortex. *J Comp Neurol* 190:597–610.
- Galaburda AM, Pandya DN (1983) The intrinsic architectonic and connectional organization of the superior temporal region of the rhesus monkey. *J Comp Neurol* 221:169–184.
- Geschwind N (1970) The organization of language and the brain. *Science* 170:940–944.

- 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.
- Ghazanfar AA, Schroeder CE (2006) Is neocortex essentially multisensory? *Trends Cogn Sci* 10:278–285.
- Giard MH, Peronnet F (1999) Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J Cogn Neurosci* 11:473–490.
- Grant KW, Seitz PF (2000) The use of visible speech cues for improving auditory detection of spoken sentences. *J Acoust Soc Am* 108:1197–1208.
- Graybiel AM (1972) Some ascending connections of the pulvinar and nucleus lateralis posterior of the thalamus in the cat. *Brain Res* 44:99–125.
- Hackett TA, Stepniewska I, Kaas JH (1998a) Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *J Comp Neurol* 394:475–495.
- Hackett TA, Stepniewska I, Kaas JH (1998b) Thalamocortical connections of the parabelt auditory cortex in macaque monkeys. *J Comp Neurol* 400:271–286.
- Hackett TA, Stepniewska I, Kaas JH (1999) Prefrontal connections of the parabelt auditory cortex in macaque monkeys. *Brain Res* 817:45–58.
- 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.
- Hackett TA (2003) The comparative anatomy of the primate auditory cortex. In: *Primate audition: Ethology and neurobiology* (Ghazanfar AA, ed), pp 199–219. Boca Raton: CRC Press.
- Hall DA, Barrett DJ, Akeroyd MA, Summerfield AQ (2005) Cortical representations of temporal structure in sound. *J Neurophysiol* 94:3181–3191.
- Hickok G, Poeppel D (2004) Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92:67–99.
- Hikosaka K, Iwai E, Saito H, Tanaka K (1988) Polysensory properties of neurons in the anterior bank of the caudal superior temporal sulcus of the macaque monkey. *J Neurophysiol* 60:1615–1637.
- Hotelling H (1933) Analysis of a complex of statistical variables into principal components. *J Educ Psychol* 24:417–520.
- Howard MA, Volkov IO, Mirsky R, Garell PC, Noh MD, Granner M, Damasio H, Steinschneider M, Reale RA, Hind JE, Brugge JF (2000) Auditory cortex on the posterior superior temporal gyrus of human cerebral cortex. *J Comp Neurol* 416:76–92.
- Jones EG (1985) *The thalamus*. New York: Plenum Press.
- Kaas JH, Hackett TA (1998) Subdivisions of auditory cortex and levels of processing in primates. *Audiol Neurootol* 3:73–85.
- Kaas JH, Hackett TA (2000) Subdivisions of auditory cortex and processing streams in primates. *Proc Natl Acad Sci U S A* 97:11793–11799.
- Kaas JH, Collins CE (2004) The resurrection of multisensory cortex in primates: connection patterns that integrate modalities. In: *The handbook of multisensory processing* (Calvert GA et al., eds), pp 285–293. Cambridge: MIT Press.
- Karbe H, Herholz K, Weber-Luxemburger G, Ghaemi M, Heiss WD (1998) Cerebral networks and functional brain asymmetry: evidence from regional metabolic changes during word repetition. *Brain Lang* 63:108–121.
- Karnath HO (2001) New insights into the functions of the superior temporal cortex. *Nat Rev Neurosci* 2:568–576.
- Kayser C, Petkov CI, Augath M, Logothetis NK (2005) Integration of touch and sound in auditory cortex. *Neuron* 48:373–384.
- Klucharev V, Mottonen R, Sams M (2003) Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Brain Res Cogn Brain Res* 18: 65–75.
- Kuriki S, Okita Y, Hirata Y (1995) Source analysis of magnetic field responses from the human auditory cortex elicited by short speech sounds. *Exp Brain Res* 104:144–152.
- Law SK, Nunez PL, Wijesinghe RS (1993) High-resolution EEG using spline generated surface Laplacians on spherical and ellipsoidal surfaces. *IEEE Trans Biomed Eng* 40:145–153.
- LeDoux JE, Ruggiero DA, Forest R, Stornetta R, Reis DJ (1987) Topographic organization of convergent projections to the thalamus from the inferior colliculus and spinal cord in the rat. *J Comp Neurol* 264:123–146.
- Lesser RP, Luders H, Morris HH, Dinner DS, Klem G, Hahn J, Harrison M (1986) Electrical stimulation of Wernicke's area interferes with comprehension. *Neurology* 36:658–663.
- Liberman AM, Cooper FS, Shankweiler DS, Studdert Kennedy M (1967) Perception of the speech code. *Psychol Rev* 74:431–461.
- Liegeois-Chauvel C, Musolino A, Chauvel P (1991) Localization of the primary auditory area in man. *Brain* 114:139–151.
- MacSweeney M, Campbell R, Calvert GA, McGuire PK, David AS, Suckling J, Andrew C, Woll B, Brammer MJ (2001) Dispersed activation in the left temporal cortex for speech-reading in congenitally deaf people. *Proc Biol Sci* 268:451–457.
- MacSweeney M, Calvert GA, Campbell R, McGuire PK, David AS, Williams SC, Woll B, Brammer MJ (2002) Speechreading circuits in people born deaf. *Neuropsychologia* 40:801–807.
- Massaro D (1998) *Perceiving talking faces. From speech perception to a behavioral principle*. Cambridge: MIT Press.
- Massaro DW (2004) From multisensory integration to talking heads and language learning. In: *The handbook of multisensory processes* (Calvert GA et al., eds), pp 153–176. Cambridge: MIT Press.
- Mitzdorf U (1991) Physiological sources of evoked potentials. *Electroencephalogr Clin Neurophysiol Suppl* 42:47–57.
- Mitzdorf U (1994) Properties of cortical generators of event-related potentials. *Pharmacopsychiatry* 27:49–51.
- Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ (2002) Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain Res Cogn Brain Res* 14:115–128.
- Molinari M, Dellanna ME, Rausell E, Leggio MG, Hashikawa T, Jones EG (1995) Auditory thalamocortical pathways defined in monkeys by calcium-binding protein immunoreactivity. *J Comp Neurol* 362:171–194.
- Mottonen R, Krause CM, Tiippana K, Sams M (2002) Processing of changes in visual speech in the human auditory cortex. *Brain Res Cogn Brain Res* 13:417–425.
- Munhall KG, Vatikiotis-Bateson E (2004) Spatial and temporal constraints on audiovisual speech perception. In: *The handbook of multisensory processing* (Calvert GA et al., eds), pp 177–188. Cambridge: MIT Press.
- Murray MM, Molholm S, Michel CM, Heslenfeld DJ, Ritter W, Javitt DC, Schroeder CE, Foxe JJ (2005) Grabbing your ear: rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cereb Cortex* 15:963–974.
- Naeser MA, Gaddie A, Palumbo CL, Stiasny-Eder D (1990) Late recovery of auditory comprehension in global aphasia. Improved recovery observed with subcortical temporal isthmus lesion vs. Wernicke's cortical area lesion. *Arch Neurol* 47:425–432.
- Nascimento-Silva S, Schroeder CE, Hackett TA, Ulbert G, Karmos JF, Smiley JF (2005) Projections from visual area V2 to auditory area Tpt in the superior temporal gyrus. In: *Society for Neuroscience*.
- Norris D, Wise R (2000) The study of prelexical and lexical processes in comprehension: Psycholinguistics and functional neuroimaging. In: *The new cognitive neuroscience* (Gazzaniga MS, ed), pp 867–880. Cambridge, MA: MIT Press.
- Nunez P (1981) *Electric fields of the brain*. New York: Oxford University Press.

- Nunez PL, Pilgreen KL (1991) The Spline-Laplacian in clinical neurophysiology: a method to improve EEG spatial resolution. *J Clin Neurophysiol* 8:397–413.
- Nunez PL, Westdorp AF (1994) The surface Laplacian, high resolution EEG and controversies. *Brain Topogr* 6:221–226.
- Nunez PL, Srinivasan R (2006) *Electric fields of the brain: the neurophysics of EEG*. New York: Oxford University Press.
- Pekkola J, Ojanen V, Autti T, Jaaskelainen IP, Mottonen R, Tarkiainen A, Sams M (2005) Primary auditory cortex activation by visual speech: an fMRI study at 3 T. *Neuroreport* 16:125–128.
- Perrin F, Pernier J, Bertrand O, Giard MH, Echallier JF (1987) Mapping of scalp potentials by surface spline interpolation. *Electroencephalogr Clin Neurophysiol* 66:75–81.
- Petrides M, Pandya DN (1988) Association fiber pathways to the frontal cortex from the superior temporal region in the rhesus monkey. *J Comp Neurol* 273:52–66.
- Poeppel D, Hickok G (2004) Towards a new functional anatomy of language. *Cognition* 92:1–12.
- Puce A, Allison T, Bentin S, Gore JC, McCarthy G (1998) Temporal cortex activation in humans viewing eye and mouth movements. *J Neurosci* 18:2188–2199.
- Raij T, Jousmaki V (2004) MEG studies of cross-modal integration and plasticity. In: *The handbook of multisensory processes* (Calvert GA et al., eds), pp 515–528. Cambridge: MIT Press.
- Rauschecker JP, Tian B, Hauser M (1995) Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268:111–114.
- Rauschecker JP, Tian B, Pons T, Mishkin M (1997) Serial and parallel processing in rhesus monkey auditory cortex. *J Comp Neurol* 382:89–103.
- Rauschecker JP, Tian B (2000) Mechanisms and streams for processing of 'what' and 'where' in auditory cortex. *Proc Natl Acad Sci U S A* 97:11800–11806.
- Rauschecker JP, Tian B (2004) Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey. *J Neurophysiol* 91:2578–2589.
- Rivier F, Clarke S (1997) Cytochrome oxidase, acetylcholinesterase, and NADPH-diaphorase staining in human supratemporal and insular cortex: evidence for multiple auditory areas. *Neuroimage* 6:288–304.
- Romanski LM, Bates JF, Goldman-Rakic PS (1999a) Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. *J Comp Neurol* 403:141–157.
- Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP (1999b) Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci* 2:1131–1136.
- Sams M, Aulanko R, Hamalainen M, Hari R, Lounasmaa OV, Lu ST, Simola J (1991) Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neurosci Lett* 127:141–145.
- Schroeder CE, Lindsley RW, Specht C, Marcovici A, Smiley JF, Javitt DC (2001) Somatosensory input to auditory association cortex in the macaque monkey. *J Neurophysiol* 85:1322–1327.
- 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.
- Schroeder CE, Smiley J, Fu KG, McGinnis T, O'Connell MN, Hackett TA (2003) Anatomical mechanisms and functional implications of multisensory convergence in early cortical processing. *Int J Psychophysiol* 50:5–17.
- Selnes OA, Knopman DS, Niccum N, Rubens AB (1985) The critical role of Wernicke's area in sentence repetition. *Ann Neurol* 17:549–557.
- Seltzer B, Pandya DN (1978) Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. *Brain Res* 149:1–24.
- Seltzer B, Pandya DN (1994) Parietal, temporal, and occipital projections to cortex of the superior temporal sulcus in the rhesus monkey: a retrograde tracer study. *J Comp Neurol* 343:445–463.
- Stein BE, Meredith MA (1993) *The merging of the senses*. Cambridge, MA: MIT Press.
- Steinschneider M, Tenke CE, Schroeder CE, Javitt DC, Simpson GV, Arezzo JC, Vaughan HG (1992) Cellular generators of the cortical auditory evoked potential initial component. *Electroencephalogr Clin Neurophysiol* 84:196–200.
- Sumbly WH, Pollack I (1954) Visual contribution to speech intelligibility in noise. *J Acoust Soc Am* 26:212–215.
- Summerfield Q, McGrath M (1984) Detection and resolution of audio-visual incompatibility in the perception of vowels. *Q J Exp Psychol A* 36:51–74.
- Summerfield Q (1987) Speech perception in normal and impaired hearing. *Br Med Bull* 43:909–925.
- Summerfield AQ (1991) Visual perception of phonetic gestures. In: *Modularity and the motor theory of speech perception* (Mattingly IG, Studert-Kennedy M, eds), pp 117–137. Hillsdale, NJ: Erlbaum.
- Summerfield Q (1992) Lipreading and audio-visual speech perception. *Phil Trans Lond* 335:71–78.
- Suter CM (1970) Principal component analysis of average evoked potentials. *Exp Neurol* 29:317–327.
- Sweet RA, Dorph-Petersen KA, Lewis DA (2005) Mapping auditory core, lateral belt, and parabelt cortices in the human superior temporal gyrus. *J Comp Neurol* 491:270–289.
- Thesen T, Hansen PC, Campbell R, Osterbauer R, Calvert GA (2005) Event-related fMRI of auditory-visual speech perception. In: *Society for Cognitive Neuroscience*, Vol. 216, p Suppl S.
- van Atteveldt N, Formisano E, Goebel R, Blomert L (2004) Integration of letters and speech sounds in the human brain. *Neuron* 43:271–282.
- van Wassenhove V, Grant KW, Poeppel D (2005) Visual speech speeds up the neural processing of auditory speech. *Proc Natl Acad Sci U S A* 102:1181–1186.
- Vaughan HG Jr, Arezzo JC (1988) The neural basis of event-related potentials. In: *Human event-related potentials*, Vol. 3 (Picton TW, ed), pp 45–96. Amsterdam: Elsevier.
- von Economo C (1929) *The cytoarchitecture of the human cerebral cortex*. London: Oxford University Press.
- Wallace MN, Johnston PW, Palmer AR (2002) Histochemical identification of cortical areas in the auditory region of the human brain. *Exp Brain Res* 143:499–508.
- Wallace MT, Ramachandran R, Stein BE (2004) A revised view of sensory cortical parcellation. *Proc Natl Acad Sci U S A* 101:2167–2172.
- Ward AA, Peden JK, Sugar O (1946) Cortico-cortical connections in the monkey with special reference to area 6. *J Neurophysiol* 9:353–462.
- Wepsic JG (1966) Multimodal sensory activation of cells in the magnocellular medial geniculate nucleus. *Exp Neurol* 15:299–318.
- Wessinger CM, VanMeter J, Tian B, Van Lare J, Pekar J, Rauschecker JP (2001) Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *J Cogn Neurosci* 13:1–7.
- Westfall PH, Tobias RD, Rom D, Wolfinger RD, Hochberg Y (1999) *Multiple comparisons and multiple tests using SAS*. Cary, NC: SAS Institute Inc.
- Winer JA (1984) The human medial geniculate body. *Hear Res* 15:225–247.
- Winer JA (1992) The functional architecture of the medial geniculate body and the primary auditory cortex. In: *The mammalian auditory pathway* (Webster DB et al., eds), pp 222–409. New York: Springer-Verlag.
- Wright TM, Pelphey KA, Allison T, McKeown MJ, McCarthy G (2003) Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cereb Cortex* 13:1034–1043.

- Zatorre RJ, Bouffard M, Belin P (2004) Sensitivity to auditory object features in human temporal neocortex [erratum appears in J Neurosci 2004;24(21):5078]. J Neurosci 24:3637–3642.
- Zhou B, Green DM, Middlebrooks JC (1992) Characterization of external ear impulse responses using Golay codes. J Acoust Soc Am 92:1169–1171.

APPENDIX

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi: [10.1016/j.neuroscience.2006.11.036](https://doi.org/10.1016/j.neuroscience.2006.11.036).

(Accepted 8 November 2006)
(Available online 22 January 2007)