Visual activation and audiovisual interactions in the auditory cortex during speech perception: intracranial recordings in humans.

Julien Besle, Catherine Fischer, Aurélie Bidet-Caulet, Francoise Lecaignard, Olivier Bertrand, Marie-Hélène Giard

Behavioral/Systems/Cognitive

Visual Activation and Audiovisual Interactions in the Auditory Cortex during Speech Perception: Intracranial Recordings in Humans

Julien Besle,[1,2] Catherine Fischer,[1,2,3] Aurélie Bidet-Caulet,[1,2] Francoise Lecaignard,[1,2] Olivier Bertrand,[1,2] and Marie-Hélène Giard[1,2]

[1]INSERM, U821, "Brain Dynamics and Cognition," 69500 Lyon, France, [2]Universite Lyon 1, 69622 Lyon, France, and [3]Hospices Civils de Lyon, Neurological Hospital, 69003 Lyon, France

Hemodynamic studies have shown that the auditory cortex can be activated by visual lip movements and is a site of interactions between auditory and visual speech processing. However, they provide no information about the chronology and mechanisms of these cross-modal processes. We recorded intracranial event-related potentials to auditory, visual, and bimodal speech syllables from depth electrodes implanted in the temporal lobe of 10 epileptic patients (altogether 932 contacts). We found that lip movements activate secondary auditory areas, very shortly (≈10 ms) after the activation of the visual motion area MT/V5. After this putatively feedforward visual activation of the auditory cortex, audiovisual interactions took place in the secondary auditory cortex, from 30 ms after sound onset and before any activity in the polymodal areas. Audiovisual interactions in the auditory cortex, as estimated in a linear model, consisted both of a total suppression of the visual response to lipreading and a decrease of the auditory responses to the speech sound in the bimodal condition compared with unimodal conditions. These findings demonstrate that audiovisual speech integration does not respect the classical hierarchy from sensory-specific to associative cortical areas, but rather engages multiple cross-modal mechanisms at the first stages of nonprimary auditory cortex activation.

Key words: **auditory; visual; intracranial; speech; human; ERPs**

Introduction

Visual cues from lip movements can deeply influence the auditory perception of speech and, thus, the subjective experience of what is being heard (Cotton, 1935; McGurk and MacDonald, 1976). This suggests that visual information can access processing in the auditory cortex. Indeed, several hemodynamic imaging studies in humans have shown that the auditory cortex can be activated by lip reading (Paulesu et al., 2003) [including primary cortex (Calvert et al., 1997)], and is a site of interactions between auditory and visual speech cues (Skipper et al., 2007) [including primary cortex (Miller and D'Esposito, 2005)]. The results have mainly been interpreted in an orthodox model of organization of the sensory systems, in which auditory and visual cues are first processed in separate unisensory cortices, and then converge in multimodal associative areas (Mesulam, 1998). Visual influences in the auditory cortex would result from feedback projections from those polysensory areas (Calvert et al., 2000; Miller and D'Esposito, 2005), particularly from the superior temporal sulcus (STS). Indeed, this structure is known to respond to both visual and auditory inputs and has repeatedly been found to be active in functional magnetic resonance imaging (fMRI) studies of audio-visual integration of speech (Miller and D'Esposito, 2005), as well as nonspeech stimuli (Beauchamp et al., 2004).

However, the poor temporal resolution of hemodynamic imaging prevents access to the temporal dynamics of the cross-modal processes, and thus to the neurophysiological mechanisms by which visual information can influence auditory processing. Furthermore, the model of late multisensory convergence is being challenged by growing evidence that multisensory interactions can take place at early stages of processing, both in terms of anatomical organization and timing of activations (for review, see Bulkin and Groh, 2006; Ghazanfar and Schroeder, 2006). In the speech domain, several studies have shown that auditory event-related potentials (ERPs) can be altered by visual speech cues as early as the N1 stage, ~100 ms (Besle et al., 2004b; Möttönen et al., 2004; van Wassenhove et al., 2005), that is, during the building of an auditory neural representation (Näätänen and Winkler, 1999). However, because of the poor spatial resolution of this technique, these effects may actually have arisen in the STS, which is close to, and has the same spatial orientation as, the supratemporal plane subtending the auditory cortex.

In this study, we exploit both the spatial and temporal resolutions allowed by invasive electrophysiological recordings to elucidate the precise timing and mechanisms by which visual lip movements can alter the auditory processing of speech. We recorded intracranial ERPs evoked by auditory, visual, and bimodal syllables in 10 epileptic patients enrolled in a presurgical evalua-

Received June 23, 2008; revised Sept. 25, 2008; accepted Nov. 24, 2008.

We are grateful to Pierre-Emmanuel Aguera for his invaluable technical assistance.

Correspondence should be addressed to Julien Besle at the above address. E-mail: julien.besle@inserm.fr.

DOI:10.1523/JNEUROSCI.2875-08.2008

 0270-6474/08/2814301-10$15.00/0

tion program and implanted with depth electrodes, mainly in the temporal cortex (see Fig. 1). Our goals were (1) to determine whether the auditory cortex can be activated by lip movements before or after polysensory areas, and (2) to describe the precise chronology, location, and nature of the audiovisual interactions in the temporal cortex, by comparing the bimodal ERPs with the unimodal responses.

Materials and Methods

Patients. We recorded data from 10 patients (4 males; ranging from 19 to 59 years of age) suffering from pharmacologically resistant partial epilepsy, and who were candidates for surgery. Because the location of the epileptic focus could not be identified using noninvasive methods, they were stereotactically implanted with multicontact depth probes (Fig. 1). Electrophysiological recording is part of the functional evaluation that is performed routinely before epilepsy surgery in these patients. In accordance with the French regulations concerning invasive investigations with a direct individual benefit, patients were fully informed about the electrode implantation, stereotactic EEG, and evoked potential recordings, and the cortical stimulation procedures used to localize the epileptogenic and functional brain areas. All patients gave their informed consent to participate in the experiment. The signals described here were recorded away from the seizure focus. Several days before EEG recordings, antiepileptic drugs administered to the patients had been either discontinued or drastically reduced. No patient was administered with benzodiazepines. None of the patients reported any auditory or visual complaint.

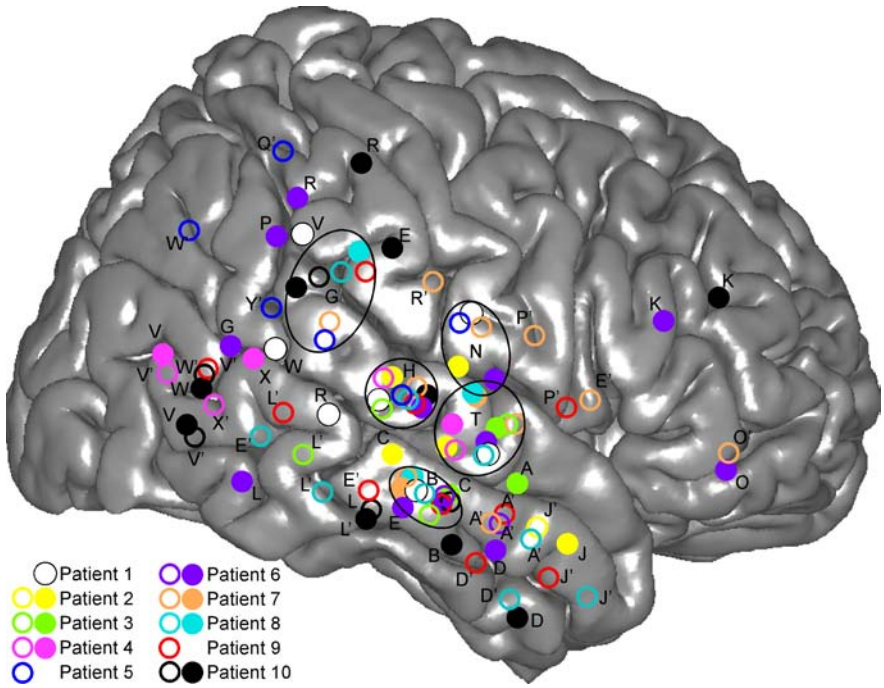

Figure 1. Location of multicontact electrodes in the 10 patients, reported on a common 3D representation of the right hemisphere. The 3D representation of the cortex has been segmented from the anatomical MRI of the right hemisphere of the standard MNI brain. Open and filled circles are electrodes implanted in the left and right hemispheres, respectively. *X* and *Y* coordinates of each electrode of each patient have been normalized to the MNI space using the Talairach method. Letters correspond to the name prefix of electrode contacts. The names of left-implanted electrodes are followed by a prime.

Stimuli and task. Stimuli and tasks were similar to those used in a previous scalp ERP study in healthy subjects and have been described in detail by Besle et al. (2004b). Four French syllables (/pa/, /pi/, /po/ and /py/, three exemplars of each) were presented in three conditions: visual (lip movements), auditory, and audiovisual. All 36 stimuli were presented in random order, over 8 blocks of 66 stimuli each. Before each block, one of the four syllables was designated as the target and the patient's task was to click on a mouse whenever he/she heard the target (auditory and audiovisual conditions). No response was required in the visual-only condition to avoid engaging unnatural and exaggerated attention to the visual modality (by trying to lip-read), compared with the other two conditions (Besle et al., 2004a). The target syllables were not used to compute the ERPs. Stimuli were selected from a wider set of audiovisual syllables uttered by the same speaker so that, in all 12 syllables, the visual part of the syllable preceded the sound onset by exactly 240 ms (six video frames). The onset of the auditory syllable (or its corresponding point in time in visual-only trials) was taken as time 0. Visual stimuli were sampled and played at 25 fps and auditory stimuli at 41.1 kHz. Contrary to our scalp ERP study, the auditory syllables were presented with headphones. The stimulation setup was otherwise the same.

EEG recording and ERP analysis. Intracranial recordings were performed at the Functional Neurology and Epilepsy Department of Lyon Neurological Hospital. EEG was recorded from 64 or 128 intracranial electrode contacts referenced to an intracranial contact away from the temporal cortex. The ground electrode was at the forehead. Signals were amplified, filtered (0.1–200 Hz bandwidth), and sampled at 1000 Hz (Synamps, Neuroscan Labs) for the first five patients, and were amplified, filtered (0.1–200 Hz bandwidth), and sampled at 512 Hz (Brain Quick SystemPLUS Micromed) for the next five patients.

All signal analyses were performed using the ELAN-Pack software developed at INSERM U821. EEG signal was digitally bandpass-filtered (0.2–100 Hz) and notch-filtered at 50 Hz. Trials including interictal discharges were excluded by an automatic procedure: for each condition of presentation, we excluded any trial including at least one time sample with an amplitude >5 SDs relative to the mean amplitude (across the trials) over a period from 300 ms before time 0 to 600 ms after. However, if a single electrode contact participated in >6% of trial rejection for a given condition, this contact was excluded. After rejection, the mean number of remaining electrode contacts per subject was 56 and the mean number of averaged trials per subject was 104, 106, and 107 for the visual, auditory, and audiovisual syllables, respectively. ERPs were averaged between −300 and 600 ms (around time 0) and baseline-corrected relative to the mean amplitude of the (−300 −150 ms) time-window (because visual ERPs emerged as soon as −100 ms). ERPs were computed both from monopolar montages (all contacts referred to the same reference site) and from bipolar montages (every contact referred to its immediately adjacent neighbor). Whereas monopolar montages allow comparison of the response polarity with those of scalp-recorded ERPs, bipolar montages emphasize the contribution of local generators. Spatiotemporal maps along the different contacts of an electrode were computed using bilinear interpolation. Only bipolar maps are shown here (Fig. 2).

Statistical analysis. For each patient, we considered only statistically significant effects, after correction for multiple tests in the temporal and spatial dimensions. Because the electrode implantation was different for each patient, the data from each patient were tested independently, using single trials as statistical units. For all statistical tests (with one exception), data were downsampled to 50 Hz, with the amplitude at each time sample being the averaged amplitude over the 40-ms time-window around the sample, to reduce the number of tests. To reduce the contribution of distant generators by volume conduction in the observed effects, we only considered differences that were statistically significant in both monopolar and bipolar data, except as otherwise mentioned.

The emergence of unimodal visual and auditory ERPs from baseline (unimodal activations) was statistically tested with paired Wilcoxon signed-rank tests between −150 and 600 ms (38 samples) at all contacts (mean number of contacts per patient: 93), which yields ~3500 tests per

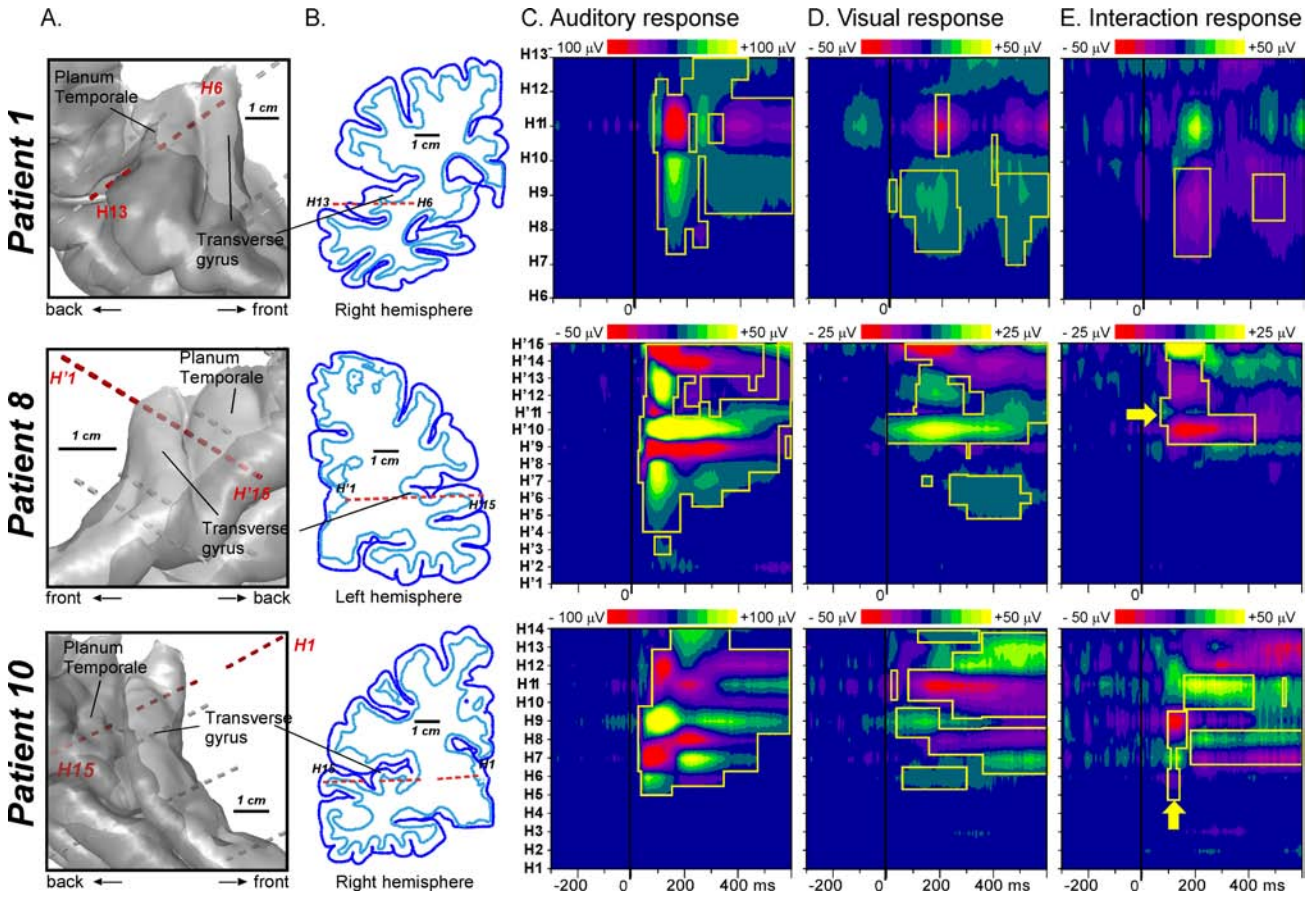

Figure 2. Auditory, visual, and interaction responses recorded from a multicontact electrode (H) passing through the transverse gyrus and the planum temporale in three patients (bipolar data). Each row corresponds to one particular patient. ***A***, Location of the multielectrode relative to a 3D rendering of the cortical surface (superior part of the temporal cortex), segmented from each patient's anatomical MRI. ***B***, Position of the multicontact electrode in the coronal plane containing the electrode. The dark blue lines delineate the cortical surface and the light blue lines the interface between the white and gray matters. ***C***, Spatiotemporal profile of the auditory response along the multicontact electrode axis from 300 ms before to 600 ms after sound onset. ***D***, Spatiotemporal profile of the visual response in the same latency range. ***E***, Spatiotemporal profile of the interaction response (computed as the difference between the bimodal response and the sum of the unimodal responses). The yellow boxes delimit the responses with statistically significant amplitudes [$p < 0.00001$ for auditory (***C***) and visual (***D***) panels; $p < 0.001$ corrected for multiple comparisons on the temporal dimension for interaction (***E***) panel]. Yellow arrows indicate the sites and times for which the interaction pattern corresponds to the modulation of the transient auditory response.

patient and condition. To correct for multiple testing, we used the Bonferroni inequality and set the statistical threshold to $p < 10^{-5}$. The emergence of the primary auditory response was tested on the original 1000 Hz-sampled data (512 Hz for the last five patients) between 10 and 40 ms after time 0, corresponding to 30 time samples (15 samples for the last five patients). This amounted to ~2800 tests per patient (1400 for the last five patients). The statistical threshold was thus set to $p < 10^{-5}$. Unimodal emergence tests were conducted on all contacts including rejected contacts to increase the spatial sampling, because we expected the response to be robust enough to resist the variability introduced by interictal spikes.

To test the existence of audiovisual interactions, we compared the ERP elicited by the audiovisual syllable with the sum of ERPs elicited by unimodal syllables between 0 and 200 ms (Besle et al., 2004a). We thus sought to test the null hypothesis that the bimodal trials were drawn from the same distribution as the distribution of the sum of auditory and visual trials. Because the unimodal distributions were independently estimated, we could not estimate the distribution of the unimodal sum. Thus, no straightforward statistical test existed for this case, and we had to devise an original test based on randomizations (Edgington, 1995). Each randomization consisted of (1) pairing at random N auditory and N visual trials (for this test we used the same number of auditory and visual trials and thus discarded the extra trials in one of the conditions), (2) computing the N corresponding sums of unimodal trials, (3) pooling together these N sums with the M audiovisual trials, (4) drawing at random M and N trials from this pool, and (5) computing the difference between the mean of these two samples. We did 10,000 such randomizations so as to obtain an estimate of the distribution of these differences under the null hypothesis. We then compared the actual difference between the bimodal ERP and the sum of unimodal ERPs to this distribution. This test was conducted between 0 and 200 ms (20 time-samples) at each contact left after trial rejection (56 per patient on average), which yielded on average 1120 tests per patient. Because we expected interaction effects to be less robust than the emergence of the unimodal responses, a Bonferroni correction would have been over-conservative because it assumes that the data for each test are independent and does not take into account the spatiotemporal correlation of EEG data (Manly et al., 1986). To increase sensitivity while taking into account multiple testing in the temporal dimension, we used a randomization procedure proposed by Blair and Karniski (1993): For each permutation of the dataset, the maximum number of consecutive significant time-samples in the entire 0–200 ms time-window was computed. Over all the permutations, we thus obtained the distribution of this maximum number under the null-hypothesis. In this distribution, the 95th percentile corresponds to the maximum number of consecutive significant samples one can obtain by chance with a risk of 5%. We required all the effects to last at least this number of samples. This procedure was run separately for each channel. For the spatial dimension, we considered the data to be independent and therefore set the statistical threshold to $p < 0.001$.

Electrode implantation, anatomical registration and normalization. Depth probes (diameter 0.8 mm) with 10 or 15 contacts each were inserted perpendicularly to the sagittal plane using Talairach's stereotactic

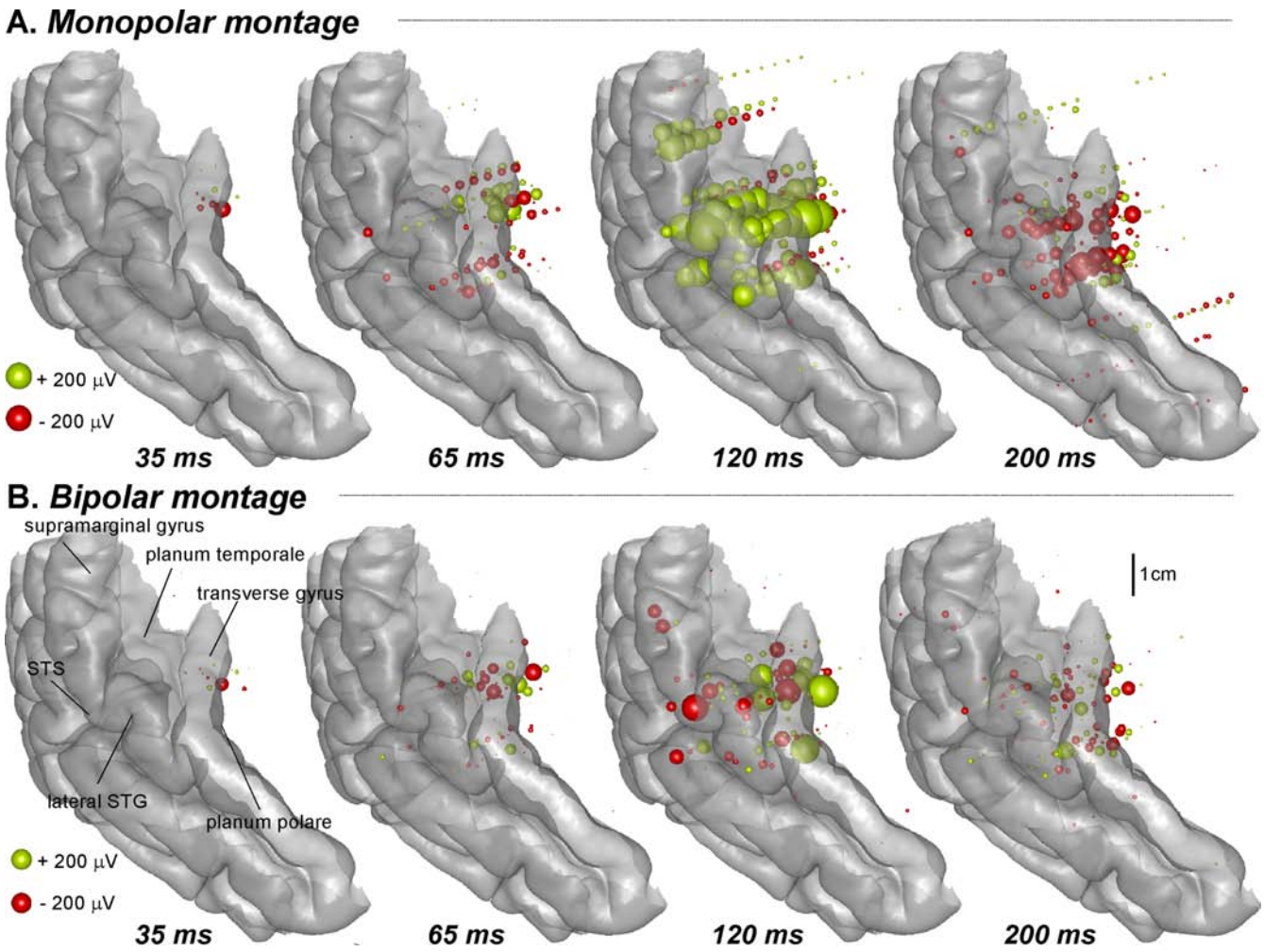

Figure 3. Statistically significant auditory response to the auditory syllables in all patients reported on a common 3D representation of the auditory cortex at 35, 65, 120 and 200 ms after the onset of the auditory stimulus. Each sphere represents the amplitude recorded at one contact in one patient. The radius of the sphere is proportional to the response amplitude, and its color indicates the polarity. Left activations were reported on the right hemisphere. ***A***, Monopolar montage (the voltage values are measured between each contact and a common reference contact) provides information about the polarity of the recorded component. ***B***, Bipolar montage (voltage values are measured between two adjacent contacts); amplitudes in this case indicate that the corresponding neural sources are very close to the site of recording.

grid (Talairach and Tournoux, 1988). Electrode contacts were 2-mm-long and spaced every 3.5 mm (center-to-center). Numbering of contacts increases from medial to lateral along an electrode track. Electrode locations were measured on x-ray images obtained in the stereotactic frame. Penetration depth for each contact was measured on the frontal x-ray image from the tip of the electrode to the midline, which was visualized angiographically by the sagittal sinus. The coregistration of the lateral x-ray image and a midsagittal MRI scan, both having the same scale of 1, allowed us to measure the electrode coordinates in the individual Talairach's space defined by the median sagittal plane, the AC–PC (anterior commissure–posterior commissure) horizontal plane and the vertical AC frontal plane, these anatomical landmarks being identified on the three-dimensional (3D) MRI scans. With this procedure, we could superpose each electrode contact onto the patients' structural MRIs. The accuracy of the registration procedure was 2 mm, as estimated on another patient's MRI obtained just after electrode explantation and in which electrode tracks were still visible. Electrode contacts locations were plotted on a 3D rendering of the temporal cortices of the patient's MRI (cortical surface segmentation by FreeSurfer software, http://surfer.nmr.mgh.harvard.edu).

All reported significant effects were localized individually on the patient's MRI. However, to facilitate qualitative comparison between patients, the locations of each patient's contacts were plotted on a common brain. The electrode coordinates of each patient were first converted from the individual Talairach's space to the normalized Talairach's space, and then to the Talairach's space of the Montreal Neurological Institute (MNI) standard brain, using quadrant linear transformation (Talairach and Tournoux, 1988). Electrode contacts and experimental effects of all patients were plotted on a 3D rendering of the right temporal lobe of the MNI standard brain. Because we saw no compelling evidence of lateralization of the responses to speech stimuli and for the sake of simplicity, left contacts were plotted on the right hemisphere by taking the opposite *x*-coordinate. Because the normalization procedure introduces localization errors attributable to anatomical variance between subjects and between hemispheres, it was only used for visualization purposes.

Results

Auditory responses

Auditory responses to the syllables were primarily recorded in the superior part of the temporal cortex. Examples of the auditory response, recorded from a multicontact electrode passing through the transverse (Heschl) gyrus and the planum temporale in three patients, are depicted in Figure 2*C*. The response consisted of a succession of transient deflections with different spatial profiles along the multicontact electrode beginning from 14 ms after the auditory syllable onset. Despite important differences in

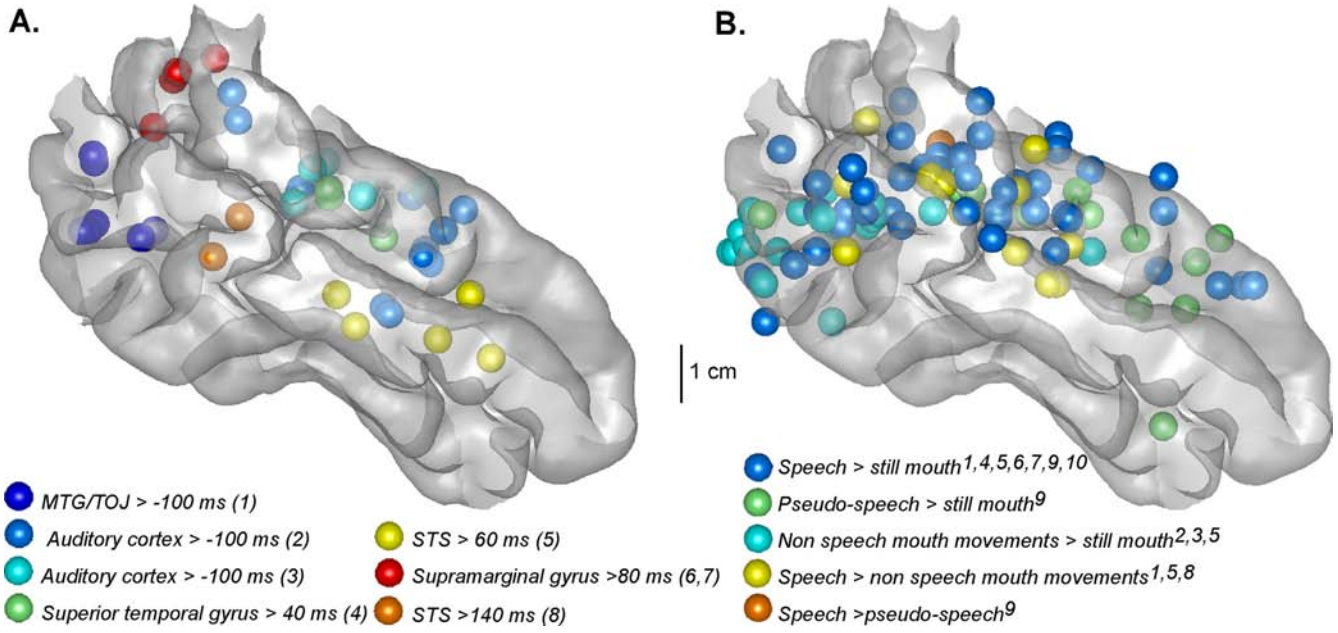

Figure 4. Comparison between (***A***) intracortically recorded responses to lip movements and (***B***) fMRI results from lipreading experiments, displayed on a lateral view of the MNI brain temporal cortex. In both cases, spheres stand for activations regardless of the size of the effect found. Left activations were reported on the right hemisphere. ***A***, Statistically significant activations evoked by lip movements alone in our experiment. Each color stands for a cluster of similar responses observed in at least two patients. Numbers in parentheses refer to the type of the visual response as characterized in supplemental Table 2, available at www.jneurosci.org as supplemental material, and described in the text. Individual coordinates have been converted into the MNI space using Talairach method. ***B***, Each color corresponds to a particular contrast that has been tested in studies by the following: (1) Calvert et al. (1997); (2) Puce et al. (1998); (3) Puce and Allison (1999); (4) MacSweeney et al. (2000); (5) Campbell et al. (2001); (6) MacSweeney et al. (2001); (7) Olson et al. (2002); (8) MacSweeney et al. (2002); (9) Paulesu et al. (2003); (10) Calvert and Campbell (2003). When originally given in Talairach space, fMRI activations were converted into the MNI space using the Talairach method.

the implantation of each patient's auditory cortex, similarities in the responses between patients are easily seen in Figure 3, where significant auditory activations from all patients are reported on a common 3D reconstruction of the right temporal cortex. The first auditory response, found in seven patients from 14 to 40 ms (mean onset, 23 ms), was circumscribed to the medial part of the transverse gyrus, which subtends the primary auditory cortex (see also supplemental Table 1, available at www.jneurosci.org as supplemental material). Auditory activation then spread to the lateral part of the transverse gyrus and to the planum temporale, and could be observed as components of either positive or negative polarity in monopolar montages (see Material and Methods) that peaked ~65 ms. Activation spread further to the superior temporal gyrus (STG), the supramarginal gyrus, and the STS and appeared as a positive component peaking ~120 ms. A component with opposite polarity but similar spatial range was recorded ~200 ms. From ~70 ms, other responses with smaller amplitudes were recorded outside the auditory cortex, but they will not be described here because precise description of the auditory response to speech is beyond the scope of this report.

Visual responses

In our stimuli, faint lip movements began 240 ms before the onset of the auditory syllable. Thus, the cortical responses to visual syllables could precede the sound onset and will be reported in this case with negative latencies. In general, the visual responses recorded in the temporal cortex were of lesser amplitude than the auditory responses described in Auditory responses. They were also generally long-lasting and did not change polarity over time (Fig. 2, compare panels *C* and *D*). Furthermore, they showed more variability between subjects, and it was not possible to ascribe individual visual responses to a few common components as was done for the auditory response. Instead, we selected statistically significant visual responses present in at least two patients, and we categorized them with respect to their latency, the cortical region in which they were recorded, and their similarity with the unimodal auditory response at the same contacts. This procedure yielded 13 different types of visual responses, whose Talairach coordinates and latencies are reported in detail in the supplemental Table 2, available at www.jneurosci.org as supplemental material. The visual responses recorded in the temporal lobe are depicted in Figure 4*A* and will be described in rough chronological order.

The earliest visual responses were recorded from 100 ms before the auditory syllable, that is, 140 ms after the visual stimulus onset, both in the posterior middle temporal gyrus (response type 1) and the superior temporal cortex (response types 2 and 3), including the lateral transverse (Heschl) gyrus, the planum temporale, the planum polare, the STG and the STS. Whereas type 1 responses were recorded at contacts that did not show any auditory response, type 2 and 3 visual responses in the superior temporal cortex showed a spatial pattern along electrode contacts that was strikingly similar with the transient unimodal auditory responses recorded on the same electrode. These visual responses were spatially similar either to a transient auditory response peaking ~65 ms (type 2) or to a transient auditory response peaking ~120 ms (type 3). This spatial similarity is illustrated for two patients (patient 1 and patient 8) in Figure 2*C*,*D*: In patient 8, the visual response at contact H'10 resembled the auditory ERP beginning ~50 ms (type 2). In patient 1, the visual response at H'9 resembled the transient auditory ERP beginning ~100 ms (type 3). Type 2 and type 3 visual responses were found in 12 nonadjacent sites in 5 patients, and in 6 nonadjacent sites in 4 patients, respectively. This spatial similarity with transient auditory responses strongly suggests that they were indeed generated in the auditory cortex and were not caused by volume conduction from

nonspecific cortices such as the STS. However, Figure 5 shows that these responses were not seen in general in the medial part of the transverse gyrus (corresponding to the primary auditory cortex): indeed, the locations of type 2 and 3 visual responses barely overlap the location of the earliest auditory response (<30 ms) in the medial transverse gyrus. A notable exception is patient 10, who showed visual activation at electrode contacts (H6–9) where the primary auditory response was recorded (Fig. 2, compare *C*, *D*).

After these activations, statistically significant responses common to auditory and visual conditions were observed around the anterior STS after 60 ms (response type 5), the posterior STS after 100 ms (response type 7), the supramarginal gyrus after 140 ms (response type 8), and the postcentral operculum and insula after 100 ms (response type 9). Other visual-specific responses were also recorded in the superior temporal cortex after 40 ms (type 4), the supramarginal gyrus after 80 ms (type 6), and in other structures outside the temporal cortex, generally after 140 ms: the insula (type 10), cingular gyrus (type 11), precentral operculum, or inferior frontal gyrus (type 12), and hippocampus and parahippocampal gyrus (type 13) (see supplemental Table 2, available at www.jneurosci.org as supplemental material).

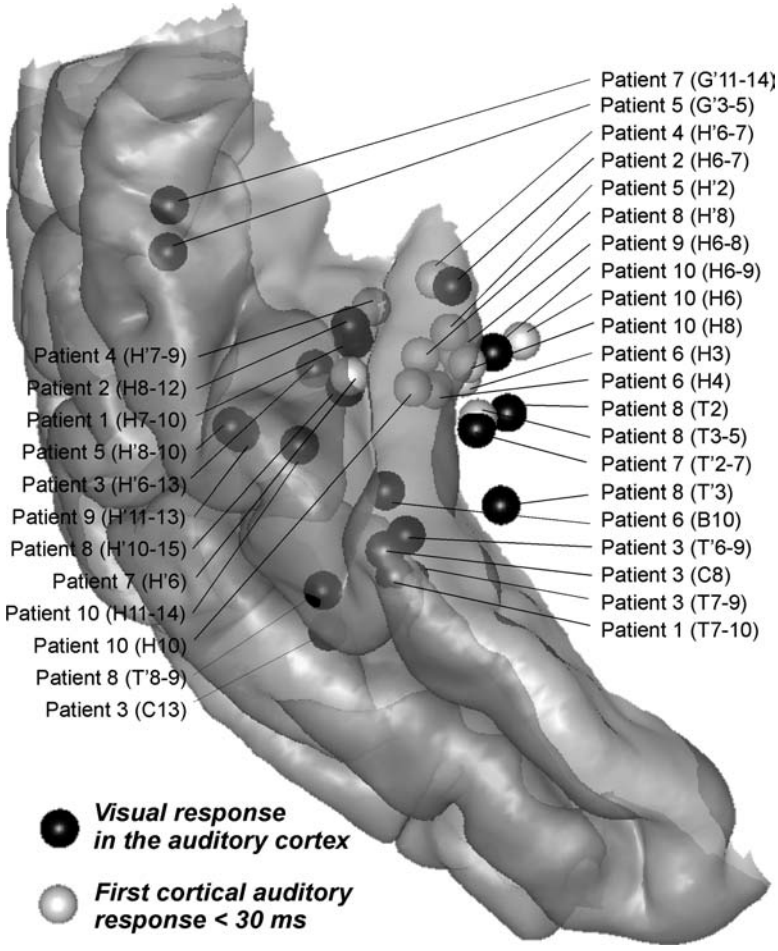

Figure 5. Comparison between the location of the visual responses recorded in the auditory cortex and the earliest auditory responses (<30 ms) recorded in the primary auditory cortex. The individual coordinates have been converted to MNI coordinates using the Talairach method and reported on the standard MNI brain. Left activations were reported on the right hemisphere.

Audiovisual interactions

Because visual responses were observed all over the superior temporal cortex at latencies preceding, or coincident with, its activation by the auditory syllables, visual and auditory speech processing should interact in this region. To estimate the latency and location of these audiovisual interactions, we subtracted the sum of unimodal ERPs from the ERP elicited by the bimodal syllable (see Materials and Methods). Statistically significant interaction responses were mainly found in the superior temporal cortex and are illustrated in Figure 2*E* for three patients (see supplemental Table 3, available at www.jneurosci.org as supplemental material, for a detailed account of all interactions responses). The spatiotemporal profile of the interaction response shows a striking similarity to the visual response, but with opposite polarity. Because the interaction and the unimodal visual response have the same amplitude order, the interaction can be interpreted as the visual response being suppressed in the bimodal condition compared with the unimodal visual condition. This pattern of interaction was found at 19 nonadjacent sites in nine patients. It could start from 30 to 160 ms after the auditory syllable onset and lasted well after 600 ms.

However, this suppression of the visual response cannot account for the entire pattern of interaction effects. As can be seen in Figure 2*E* for patients 8 and 10, parts of the significant interaction responses (indicated by arrows) do not have the same profile as the visual response. Rather, they show a spatiotemporal similarity with the transient auditory-only response, with opposite polarity (contact H′11 in patient 8, contact H9 in patient 10). Because their amplitude is smaller than that of the corresponding auditory component, they can be interpreted as a decrease of the auditory transient response in the bimodal condition compared with the auditory-only condition. This pattern of interaction was found in nine nonadjacent sites in six patients, between 40 and 200 ms after the auditory syllable onset. Note that, in four of nine nonadjacent sites, this modulation was only seen in the monopolar montage and, in four of the nine sites, we had to lower the statistical threshold (see supplemental Table 3, available at www.jneurosci.org as supplemental material). The locations of these two types of interactions are summarized in Figure 6. They were generally spatially overlapping and were often found at the same electrode for the same patient. However they were recorded outside the medial part of the transverse gyrus (except for patient 10).

Other statistically significant interactions were found but they could not be classified into categories of responses sharing spatial, temporal, or functional features (see response type 3 in supplemental Table 3, available at www.jneurosci.org as supplemental material).

Relationship between unimodal and interaction responses

Thus far, we have described each type of response separately for the entire group of patients. Given the variability of onset latencies for the same type of response between patients and the fact that not all patients showed all types of response, it was not pos-

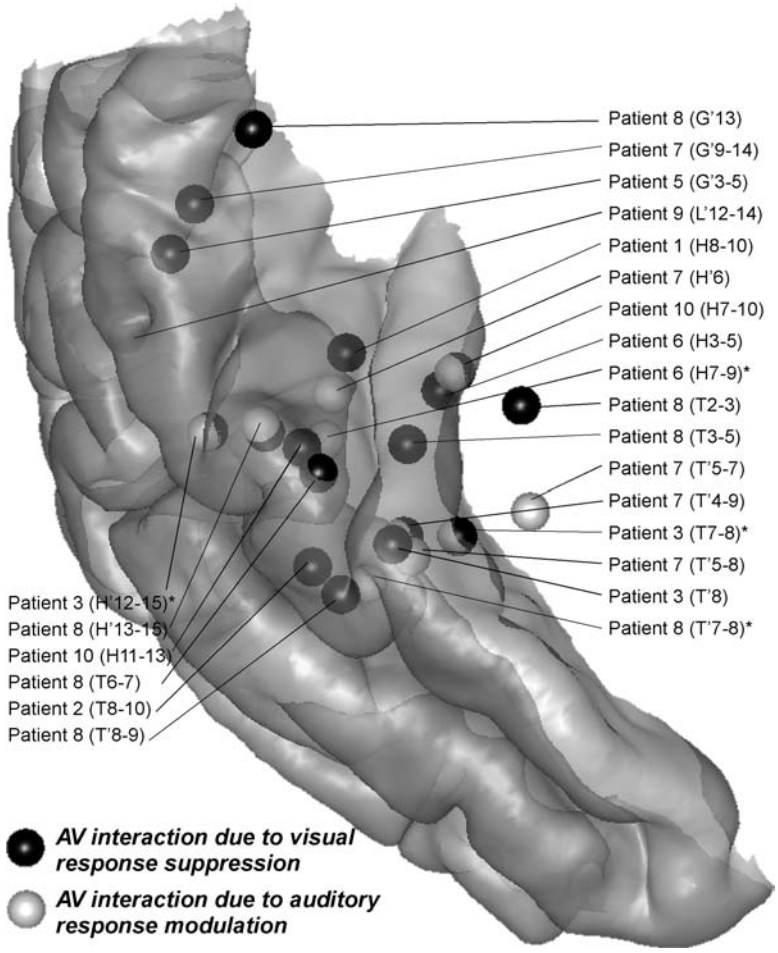

Figure 6. Locations of the interaction responses in the superior temporal cortex. The individual coordinates have been converted to MNI coordinates using the Talairach method and reported on the standard MNI brain. Left activations were reported on the right hemisphere. (*) Indicates that the effect was significant only in the monopolar data for the second type of interaction (see Results, Audiovisual interactions).

sible to compare these auditory, visual and interaction responses at the group level. However, some of the patients showed individually several types of responses. This makes the comparison of onset latencies possible and allows us to suggest, somewhat tentatively, a description of the organization of unimodal and interaction responses in the bimodal presentation condition. Table 1 presents the latencies of four types of responses: two visual responses (in the posterior middle temporal cortex and in the superior temporal cortex) and two interaction responses (modulation of the transient auditory response and suppression of the visual response) in each of the 10 patients. Only patients 8 and 10 showed all types of response, but interestingly, the succession of events was similar in the two patients: in the bimodal condition, the middle temporal cortex was first activated by the lip movements, followed 10 ms later by the auditory cortex. These two activations occurred well before the presentation of the auditory syllable. Approximately 50 ms after the presentation of the auditory syllable, the auditory transient ERPs were modulated (decreased) and this modulation was followed by the suppression of the visual response. This succession of events was respected for all the other patients (except for patient 6 in whom the suppression of the visual response began 10 ms before the modulation of the auditory response).

Behavioral results

As a group, the patients showed a shorter response time (RT) to audiovisual targets (514 ms) than to auditory target syllables (530 ms). However, this difference was only marginally significant (one-tailed Student's t test, $p = 0.069$). The patients' RTs were longer by ~100 ms compared with the RTs measured in healthy subjects (400 vs 423 ms, $p < 0.001$) with the exact same paradigm (Besle et al., 2004b). The marginal significance of the behavioral effect in patients may be attributable to the larger variability in their RTs (SD of the difference = 32 ms for epileptic patients and 22 ms for healthy subjects). In addition, the much longer RTs in patients than in healthy subjects can reflect a general slowing in executive functions, which may have blurred the perceptual enhancement at the sensory level, particularly because we have no reason to think that perception was impaired in these patients.

Discussion

This experiment brings new insights into the cortical dynamics of auditory and visual speech processing in the human brain, both in the temporal and the spatial dimensions. By recording ERPs directly from different precise locations in the temporal cortex of epileptic patients, we provide evidence that (1) lip movements can activate the secondary auditory cortex after ~150 ms of processing, and (2) auditory and visual processing of speech can interact in the secondary auditory cortex in two ways: auditory speech syllable can totally suppress the response to lip movements, and visual lip movements can modulate (slightly suppress) the transient auditory response to the syllable.

Direct feedforward visual activation of the secondary auditory cortex

Speech lip movements activated, at ~140 ms after visual stimulus onset, the temporo-occipital junction and the posterior MTG, the location of which corresponds to the movement-sensitive visual area MT/V5 in humans (d'Avossa et al., 2007). Some 10 milliseconds later, the superior temporal cortex was activated with a spatial profile very similar to that of the transient auditory components peaking ~65 or 120 ms at the same electrode. Because these transient auditory components are known to be generated in the auditory cortex (Liégeois-Chauvel et al., 1994; Yvert et al., 2005), this visual response was thus probably also generated in the auditory cortex. Those first two types of visual responses (in MT/V5 and in auditory cortex) occurred well before visual activations in other parts of the brain (at least those investigated in the study). The activation of the auditory cortex by lip movements may therefore result from a direct feedforward process. We thus replicated the auditory cortex activation by lipreading found by numerous fMRI studies, but we added the crucial information of timing to show that this cross-modal activation happens

Table 1. Begin and end latencies of the first three types of visual response and first two types of interaction response for each of the patients

	Visual response in MTG/ occipito-temporal junction (type1)		Visual response in auditory cortex (types 2 and 3)		Interaction: modulation of the auditory response		Interaction: suppression of the visual response	
Patient	Begin	End	Begin	End	Begin	End	Begin	End
Patient 1			−20	600+			110	250
Patient 2			−120	450			40	110
Patient 3			−120	600+	50	120	80	600+
Patient 4								
Patient 5			0	600+			130	600+
Patient 6	−80	350			40	120	30	600+
Patient 7			−20	450	60	200	70	600+
Patient 8	−80	400	−70	600+	50	120	70	500
Patient 9	−100	160					120	250
Patient 10	−40	600+	−30	600+	80	160	80	600+

All latencies (in ms) are measured relative to the auditory syllable onset; 600+ indicates that the response continues after 600 ms.

within a few milliseconds of activations in visual-specific areas. Although we acknowledge it is problematic to draw comparisons between humans and animals concerning such a specialized function as speech communication, recent anatomical studies have shown the existence of direct projections from the visual cortex (Hishida et al., 2003; Cappe and Barone, 2005) and from multisensory thalamic nuclei (Hackett et al., 2007) to the auditory cortex, that could account for this cross-modal activation.

There has been debate in the literature as to whether activation of the auditory cortex by lipreading occurs in the primary cortex or only in secondary auditory cortex (Calvert et al., 1997; Bernstein et al., 2002; Pekkola et al., 2005). One patient (patient 10), of seven in whom a primary auditory response could be recorded in the medial part of the transverse gyrus, showed a visual response at the same contacts. But overall, our data tend to show that this activation occurs mainly outside the primary auditory cortex, because the visual responses and the first transient auditory response were recorded at different contacts. It should, however, be kept in mind that lipreading was not the primary task of our patients, who were only asked to look at the face. It is also possible that primary auditory cortex activation by lip movements, as seen in fMRI, reflects later feedback activations from associative multimodal areas and/or other phenomena such as auditory imagery rather than genuine sensory activation. Note that our conclusion is based on negative findings, and it is still possible that the absence of visual activation in the primary auditory cortex is caused by a lack of sensitivity of our analysis.

Because lip movements were contrasted with a resting mouth, we cannot decide, on the basis of our data, whether this activation is specific to speech or to movement per se. However, Figure 4*B* presents results from 10 fMRI lipreading studies that used various contrasts: while activation by nonspeech lip movements are restricted to the posterior MTG, speech-specific lip-movement activations (speech > nonspeech lip movements contrasts) occur in the superior temporal cortex, close to the cross-modal activation found in our experiment, which suggests that this activation is indeed speech-specific.

Two forms of audiovisual interaction in the secondary auditory cortex

Once lip movements had activated the auditory cortex, audiovisual interactions in the secondary auditory cortex could take two forms. First, as soon as the secondary auditory cortex becomes activated by the auditory syllable, the long-lasting visual response seen in visual-only trials completely disappears in the audiovisual condition. Second, comparison of the spatio-temporal profiles of the interaction response with those of the unimodal auditory response reveals that some transient auditory components between 40 and 200 ms are decreased when the auditory syllable is presented with lip movements compared with when it is presented alone. These two forms of interactions may convey complementary processes of integration of speech signals in the auditory cortex that facilitate the processing of the auditory syllable.

At least two scenarios can explain the full pattern of interactions observed, depending on the type of information brought by the visual activation of the auditory cortex. One possibility is that lip movements only bring temporal information about the onset of the syllable sound. This activity would thus stop after the onset of the acoustic signal in the auditory cortex, and the subsequent modulations of the auditory components would reflect a cross-modal attentional effect by visual temporal cueing (Schwartz et al., 2004; Stekelenburg and Vroomen, 2007). Alternatively, lip movements may have brought phonetic information that "preactivate" the auditory cortex; this auditory activation resulting from the visual input would stop to leave full resources to the sensory-specific (auditory) cortex to process the acoustic/phonetic features of the syllables more efficiently. Nevertheless, the "preprocessing" of the visual syllable would result in engaging less auditory resources from the auditory cortex (response decreases) to process the syllable (Besle et al., 2004a, Giard and Peronnet, 1999). In both cases, the suppression of visual activity might reflect a ceiling effect: the auditory cortex would be maximally activated by speech sounds, and there would be no possibility for an additional activation by lip movements once the sound starts being processed. If this is the case, only the subsequent decrease in transient auditory activity might be related to audiovisual integration per se.

It is to be noted that the reduction of activity in the auditory cortex for bimodal speech is similar to the effect found on the auditory N1 component of scalp ERPs in normal subjects (Besle et al., 2004b; van Wassenhove et al., 2005), which has been associated with a behavioral gain (shorter response time) to identify the syllable in the audiovisual condition. Indeed, the modulation of the positive auditory component peaking ~120 ms is the analog of the auditory N1 decrease found in scalp ERPs using the exact same experimental paradigm (Besle et al., 2004b). This component is recorded with a positive polarity in intracranial data, because electrodes are located under the supratemporal cortical surface (Godey et al., 2001; Yvert et al., 2005). Modulation of earlier auditory components and visual responses in auditory cortex were not observed on the scalp, probably because they are

of smaller amplitudes and/or vary in polarity, and therefore cancel out at the surface. Note that the decrease of auditory activity was significant only in the monopolar data for half of the sites where it was recorded, indicating that the effect could arise from more distant parts of the cortex. But, given the latency of the effects and their spatiotemporal similarity with the transient auditory responses analogous to the P50 and N1 scalp components, they probably originate from the associative auditory cortex.

Finally, the two forms of interactions in the auditory cortex found in the present study are suppression effects in which the absolute activation in the audiovisual condition is less than the sum of auditory and visual activities. This contrasts with the principles of multisensory integration, as first enunciated by Stein and Meredith (1993) at the single neuron level, according to which successful multisensory integration results in a multiplicative increase of neural activity. Several factors can account for this discrepancy, such as that these principles were established at the cellular level on spike data, with near-threshold stimuli, while we recorded the resultant of postsynaptic activity at the population level and used suprathreshold stimuli. Furthermore, suppressive interaction effects have repeatedly been evidenced at the neural population level, even when the behavioral outcome of multimodality is a facilitation (Giard and Peronnet, 1999; Beauchamp et al., 2004). Several studies have reported that, in bimodal asynchronous stimuli, the delay separating nonauditory inputs from auditory inputs can influence the pattern of interactions seen in the auditory cortex (Ghazanfar et al., 2005; Lakatos et al., 2007), especially that long delays (>100 ms, like in our stimuli) favor suppression effects over enhancements (Ghazanfar et al., 2005). One mechanism proposed to explain this effect is that visual speech movements entrain or reset ongoing oscillations in the auditory cortex, and that the outcome of the audiovisual interaction depends on the phase of those oscillations at the speech sound onset, and therefore, on the delay between the visual and the auditory inputs (Schroeder et al., 2008). In this report, we limited our analysis and our conclusions to the evoked activity, and analysis of oscillatory activities will be reported elsewhere.

References

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.

Bernstein LE, Auer ET Jr, Moore JK, Ponton CW, Don M, Singh M (2002) Visual speech perception without primary auditory cortex activation. Neuroreport 13:311–315.

Besle J, Fort A, Giard MH (2004a) Interest and validity of the additive model in electrophysiological studies of multisensory interactions. Cogn Process 5:189–192.

Besle J, Fort A, Delpuech C, Giard MH (2004b) Bimodal speech: Early suppressive visual effects in the human auditory cortex. Eur J Neurosci 20:2225–2234.

Blair RC, Karniski W (1993) An alternative method for significance testing of waveform difference potentials. Psychophysiology 30:518–524.

Bulkin DA, Groh JM (2006) Seeing sounds: visual and auditory interactions in the brain. Curr Opin Neurobiol 16:415–419.

Calvert GA, Campbell R (2003) Reading speech from still and moving faces: the neural substrates of visible speech. J Cogn Neurosci 15:57–70.

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, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr Biol 10:649–657.

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.

Cappe C, Barone P (2005) Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. Eur J Neurosci 22:2886–2902.

Cotton JC (1935) Normal "visual hearing". Science 82:592–593.

d'Avossa G, Tosetti M, Crespi S, Biagi L, Burr DC, Morrone MC (2007) Spatiotopic selectivity of BOLD responses to visual motion in human area MT. Nat Neurosci 10:249–255.

Edgington ES (1995) Randomization tests, Ed 2: revised and expanded. New York: Marcel Dekker.

Ghazanfar AA, Schroeder CE (2006) Is neocortex essentially multisensory? Trends Cogn Sci 10:278–285.

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.

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.

Godey B, Schwartz D, de Graaf JB, Chauvel P, Liégeois-Chauvel C (2001) Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: a comparison of data in the same patients. Clin Neurophysiol 112:1850–1859.

Hackett TA, De La Mothe LA, Ulbert I, Karmos G, Smiley J, Schroeder CE (2007) Multisensory convergence in auditory cortex, II. Thalamocortical connections of the caudal superior temporal plane. J Comp Neurol 502:924–952.

Hishida R, Hoshino K, Kudoh M, Norita M, Shibuki K (2003) Anisotropic functional connections between the auditory cortex and area 18a in rat cerebral slices. Neurosci Res 46:171–182.

Lakatos P, Chen CM, O'Connell MN, Mills A, Schroeder CE (2007) Neural oscillations and multisensory integration in primary auditory cortex. Neuron 53:279–292.

Liégeois-Chauvel C, Musolino A, Badier JM, Marquis P, Chauvel P (1994) Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components. Electroencephalogr Clin Neurophysiol 92:204–214.

MacSweeney M, Amaro E, Calvert GA, Campbell R, David AS, McGuire P, Williams SC, Woll B, Brammer MJ (2000) Silent speechreading in the absence of scanner noise: an event-related fMRI study. Neuroreport 11:1729–1733.

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.

Manly BJF, McAlevey L, Stevens D (1986) A randomization procedure for comparing group means on multiple measurements. Br J Math Stat Psychol 39:183–189.

McGurk H, MacDonald J (1976) Hearing lips and seeing voices. Nature 264:746–748.

Mesulam MM (1998) From sensation to cognition. Brain 121:1013–1052.

Miller LM, D'Esposito M (2005) Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. J Neurosci 25:5884–5893.

Möttönen R, Schürmann M, Sams M (2004) Time course of multisensory interactions during audiovisual speech perception in humans: a magnetoencephalographic study. Neurosci Lett 363:112–115.

Näätänen R, Winkler I (1999) The concept of auditory stimulus representation in cognitive neuroscience. Psychological Bulletin 125:826–859.

Olson IR, Gatenby JC, Gore JC (2002) A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. Brain Res Cogn Brain Res 14:129–138.

Paulesu E, Perani D, Blasi V, Silani G, Borghese NA, De Giovanni U, Sensolo S, Fazio F (2003) A functional-anatomical model for lipreading. J Neurophysiol 90:2005–2013.

Pekkola J, Ojanen V, Autti T, Jääskeläinen IP, Möttönen R, Tarkiainen A, Sams M (2005) Primary auditory cortex activation by visual speech: an fMRI study at 3 T. Neuroreport 16:125–128.

Puce A, Allison T (1999) Differential processing of mobile and static faces by temporal cortex. Neuroimage 6:S801.

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.

Schroeder CE, Lakatos P, Kajikawa Y, Partan S, Puce A (2008) Neuronal oscillations and visual amplification of speech. Trends Cogn Sci 12:106–113.

Schwartz JL, Berthommier F, Savariaux C (2004) Seeing to hear better: evidence for early audio-visual interactions in speech identification. Cognition 93:B69–B78.

Skipper JI, Goldin-Meadow S, Nusbaum HC, Small SL (2007) Speech-associated gestures, Broca's area, and the human mirror system. Brain Lang 101:260–277.

Stein BE, Meredith MA (1993) The merging of the senses, Ed 1. Cambridge, MA: MIT.

Stekelenburg JJ, Vroomen J (2007) Neural correlates of multisensory integration of ecologically valid audiovisual events. J Cogn Neurosci 19:1964–1973.

Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. New York: Thieme Medical.

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.

Yvert B, Fischer C, Bertrand O, Pernier J (2005) Localization of human supratemporal auditory areas from intracerebral auditory evoked potentials using distributed source models. Neuroimage 28:140–153.