NIH Public Access
Author Manuscript
Neuroimage. Author manuscript; available in PMC 2009 January 1.

Published in final edited form as:
Neuroimage. 2008 January 1; 39(1): 423–435.

Spatio-temporal Dynamics of Audiovisual Speech Processing

Lynne E. Bernstein[a,*], **Edward T. Auer Jr.**[b], **Michael Wagner**[c], and **Curtis W. Ponton**[d]

a *Communication Neuroscience Department, House Ear Institute, 2100 W. Third St., Los Angeles, CA 90057 USA.*

b *Department of Speech, Language, and Hearing, University of Kansas, Lawrence, KS, USA.*

c *Compumedics Neuroscan, Heussweg 25, 20255 Hamburg, Germany.*

d *Compumedics Neuroscan USA, Ltd, 6605 W W.T. Harris Blvd, Charlotte, NC 28269 USA.*

Abstract

The cortical processing of auditory-alone, visual-alone, and audiovisual speech information is temporally and spatially distributed, and functional magnetic resonance imaging (fMRI) cannot adequately resolve its temporal dynamics. In order to investigate a hypothesized spatio-temporal organization for audiovisual speech processing circuits, event-related potentials (ERPs) were recorded using electroencephalography (EEG). Stimuli were congruent audiovisual /bα/, incongruent auditory /bα/ synchronized with visual /gα/, auditory-only /bα/, and visual-only /bα/ and /gα/. Current density reconstructions (CDRs) of the ERP data were computed across the latency interval of 50-250 milliseconds. The CDRs demonstrated complex spatio-temporal activation patterns that differed across stimulus conditions. The hypothesized circuit that was investigated here comprised initial integration of audiovisual speech by the middle superior temporal sulcus (STS), followed by recruitment of the intraparietal sulcus (IPS), followed by activation of Broca's area (Miller and d'Esposito, 2005). The importance of spatio-temporally sensitive measures in evaluating processing pathways was demonstrated. Results showed, strikingly, early (< 100 msec) and simultaneous activations in areas of the supramarginal and angular gyrus (SMG/AG), the IPS, the inferior frontal gyrus, and the dorsolateral prefrontal cortex. Also, emergent left hemisphere SMG/AG activation, not predicted based on the unisensory stimulus conditions was observed at approximately 160 to 220 msec. The STS was neither the earliest nor most prominent activation site, although it is frequently considered the *sine qua non* of audiovisual speech integration. As discussed here, the relatively late activity of the SMG/AG solely under audiovisual conditions is a possible candidate audiovisual speech integration response.

Introduction

Talkers produce both optical and acoustic phonetic signals. *Phonetic* refers to the physical aspects of speech signals that encode the consonants and vowels (as well as prosody) of a language. When a talker can be seen as well as heard, perceivers typically integrate phonetic attributes of both the optical and acoustic stimuli. Integration has been demonstrated in the laboratory using many different tasks. For example, being able to see as well as hear a talker results in substantial gains to comprehending speech in noise (MacLeod and Summerfield, 1987; Sumby and Pollack, 1954), improvements in comprehending difficult messages under

*Corresponding author: Lynne E. Bernstein Communication Neuroscience Department House Ear Institute, Los Angeles, CA 900057 Email: lbernstein@hei.org Phone: 213 353 7044 Fax: 213 413 0950

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

good listening conditions (Arnold and Hill, 2001; Reisberg et al., 1987), detecting speech under adverse signal-to-noise conditions (Bernstein et al., 2004b; Grant, 2001; Grant and Seitz, 2000), compensation for filtering out various acoustic frequency bands (Grant and Walden, 1996) or due to hearing loss (Erber, 1975; Grant et al., 1998), and super-additive levels of speech identification from stimulus combinations of extremely minimal auditory and visible speech information (Breeuwer and Plomp, 1986; Iverson et al., 1998; Kishon-Rabin et al., 1996; Moody-Antonio et al., 2005).

The best known audiovisual speech integration example is the McGurk effect (McGurk and MacDonald, 1976). The McGurk effect is said to have occurred when, for example, an auditory stimulus “ba” is paired with a visual stimulus “ga,” and the perceiver reports that the talker said “da.” Because the perceptual effect is sub-phonemic (i.e., a change in the perceived place of articulation speech feature), the effect has reasonably been attributed to phonetic information integration. That is, the effect is considered to be primarily perceptual and not due to higher-level linguistic processing such as lexical.

Neuroimaging studies have used mismatched audiovisual speech to investigate mechanisms of audiovisual phonetic integration. However, not all effects obtained with audiovisual speech can be attributed to processing the phonetic information in the stimuli, because natural speech stimuli afford multiple attributes (e.g., the talker's face, gender, low-level visual features, voice timbre, loudness, etc.). that can engage a broad range of perceptual and cognitive processes Use of stimuli that engage higher levels of psycholinguistic processing (e.g., spoken words or sentences) complicate interpretation of results (Calvert et al., 1999; Calvert et al., 2000). Therefore, isolation of audiovisual phonetic integration effects requires phonetic level stimulus control and variation (Besle et al., 2004; Colin et al., 2004; Colin et al., 2002; Jones and Callan, 2003; Miller and d'Esposito, 2005), which can be accomplished by using nonsense syllables that are without semantic content.

In a functional magnetic resonance imaging experiment (fMRI), greater activity within the right supramarginal gyrus (SMG) and left inferior parietal lobule (Jones and Callan, 2003) was obtained with an incongruent nonsense syllable stimulus versus a congruent syllable. The left intraparietal sulcus (IPS) and parietal lobules were differentially sensitive to congruent and incongruent audiovisual vowels in another fMRI study (Saito et al., 2005). In a study in which subjects reported whether asynchronous audiovisual nonsense syllable stimuli were fused or not (Miller and d'Esposito, 2005), areas that were sensitive to speech fusion were Heschl's gyrus, the middle superior temporal sulcus (STS), the middle IPS, and the inferior frontal (IF) gyrus (Broca's area).

On the basis of their results from the fusion task and the literature, Miller and d'Esposito (2005) hypothesized an integration pathway for audiovisual speech. According to their hypothesis, “the middle STS is the core of the perceptual fusion network, a region where auditory and visual modalities are first combined for the purposes of identification. The intelligible speech pathway starts here and progresses anteriorly along the STS/superior temporal gyrus. … In the case of imperfect correspondence of auditory and visual signals, the IPS is recruited by the STS to effect spatiotemporal transformations … and accumulate evidence toward achieving a match. Broca's area would then be recruited only in instances in which greater effort is required … to parse speech into intelligible components” (p. 5891-2). This pathway is plausible, given that these regions are among ones whose activity is replicated across audiovisual speech and face-voice experiments (Beauchamp et al., 2004; Callan et al., 2003; Calvert et al., 2000), and given that anatomical studies have shown connections among these areas (Pandya and Yeterian, 1996). But the hypothesized pathway has a temporal dimension to its specification, and temporal dynamics on the scale that is relevant here (approximately 10s of milliseconds) are not adequately resolved using the fMRI methodology,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

which measures the relatively slow blood oxygen-level dependent (BOLD) signal (Buckner, 2003). The study presented here used electroencephalographic (EEG) event-related potentials (ERPs) to investigate spatio-temporal dynamics of audiovisual speech processing and to investigate the circuit hypothesized in Miller and d'Esposito (2005).

Models based on ERPs are appropriate for investigating hypotheses with both temporal and spatial attributes, with the caveat that the spatial dimension is relatively broad. Here, scalp-recorded ERPs were obtained for AV congruent (AVc /bα/) and AV incongruent (AVi auditory /bα/ and visual /gα/) stimuli, as well as for auditory-only (AO) /bα/ and /dα/, and visual-only (VO) /bα/ and /gα/. Cortical source generator localization using EEG can be achieved with modest spatial resolution using modeling methods such as current density reconstruction (CDR) (Fuchs et al., 1999), which was used here to achieve resolution of approximately 1 to 2 cm. The CDRs were computed on every millisecond of ERP data that had been low-pass filtered at 70Hz, thus resolving events at the same resolution as the underlying ERPs due to the linearity of the CDR computations. Given the low-pass filtering, the ERPs in this study were sensitive to events at the resolution of 7 msec, according to the sampling theorem.

CDR models spatio-temporal cortical response patterns using a large number of distributed dipole sources, without making assumptions regarding the number or dynamical properties of the dipoles. The CDR models here were used to evaluate the hypothesized spatio-temporal circuit in Miller and d'Esposito (2005). Support was obtained for the relevance to audiovisual speech processing of the hypothesized regions of activity but not for the hypothesized temporal dynamics.

Materials and Methods

Subjects

Twelve right-handed subjects (mean age 30, range 20 to 37 years) were recruited who had previously participated in a screening test of the stimuli and had demonstrated their susceptibility to the McGurk effect (McGurk and MacDonald, 1976). In the screening test, 48 stimuli were presented that combined a visual token of “tha,” “ga,” “ba,” and “da,” and an auditory token from each of the same tokens. For the classic McGurk stimulus with auditory “ba” and visual “ga,” all the subjects responded with a non-“ba” response on 50% or more of the trials (mean non-“ba” response of 90%).

All participants were neurologically normal (no reported head injuries resulting in a loss of consciousness), with normal pure-tone thresholds. Prior to testing, the purpose of the study was explained to each subject and informed and written consent was obtained. Subjects were paid for their participation.

Stimuli

Natural productions of the AV stimuli /bα/ and /gα/ were video-recorded at a rate of 29.97 frames/sec. (An acoustic /dα/ stimulus was also presented, but the results are not reported here.) Stimulus tokens were selected so that the video end frames could be seamlessly dubbed to the beginning frames of either the same stimulus (e.g., /ba/ to /bα/) or the alternate stimulus (e.g., / gα/ to / bα/). Thus, the transition from stimulus to stimulus would not result in an evoked response. In addition, the video clips were edited by removing still frames leading into the stimulus and leading out of it. Alternate frames were also removed from the quasi-steady state portion of the vowel, resulting in a total of 20 video frames (599 ms) per stimulus. The acoustic / bα/ token was 450 ms in duration.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Figure 1 shows two measures of the video stimulus dynamics in relationship to the acoustic signal. Individual video frames were used to measure in pixels the interlip distance (mouth opening) and the jaw drop. The vertical lines in the figure show the onset and offset of the acoustic stimulus. The face dynamics were somewhat different across stimuli, as would be expected, given that they were different phonemes.

For the congruent AV / bα/ stimulus (AVc), the natural relationship between the visible speech movement and the auditory speech was maintained. Therefore, the visual movement began before the onset of sound, that is, partway through the 7th video frame. For the incongruent combination of auditory /bα/ and visual /gα/ (AVi), the acoustic / bα/ signal was dubbed so that its onset corresponded with the onset of the original acoustic /gα/ for that token. In order to guarantee the audio-visual synchrony of the stimuli, the stimuli were dubbed to video tape using an industrial betacam SP video tape deck, thus locking their temporal relationships. The audio was amplified through a Crown amplifier for presentation via earbuds. In order to guarantee synchrony for data averaging of the EEG, a custom trigger circuit was used to insert triggers from the video tape directly into the Scan™ acquisition system.

For the VO /bα/ and /gα/ condition, the auditory portion of the stimuli was muted. For the AO condition, the video was turned off. Also, in the AO condition, an oddball /dα/ stimulus was presented, but none of the responses to oddball stimuli were reported for this study (see below).

Procedure

Subjects were tested while seated comfortably in an electrically shielded and sound-attenuated booth. All of the EEG recordings were obtained on a single day. The data were collected during a mismatch negativity paradigm in which standards were presented in 87% of trials pseudo-randomly ordered with 13% of deviant trials. Each stimulus was tested as both a standard and a deviant. For example, auditory /bα/ occurred as a standard versus auditory /da/ as a deviant, and vice versa in another run. Thus there were two runs for each condition. 2200 trials were presented per subject per condition. However, for this study, *only* the trials that contained *standard* stimuli were analyzed. That is, the/bα/ trials during AO condition with /dα/ as standard were not analyzed for this study. During the auditory-only stimulus presentations, subjects watched a video of a movie (entertainment—not the visual syllable stimuli) with the sound off. Visual stimuli were viewed at a distance of 1.9 m from the screen. Throughout the experiment, subjects were not required to respond behaviorally to the stimuli. Testing took approximately 4.5 hours per subject, and subjects were given rests between runs.

Electrophysiological recordings

Thirty silver/silver-chloride electrodes were placed on the scalp using a conductive water-soluble paste at locations based on the International 10/20 recording system (Jasper, 1958). A reference electrode was placed on the forehead at Fpz, with a ground electrode located 2 cm to the right and 2 cm up from Fpz. Vertical and horizontal eye movements were monitored on two differential recording channels. Electrodes located above and below the right eye were used to monitor vertical eye movements. Horizontal eye movements were recorded by a pair of electrodes located on the outer canthus of each eye. For each stimulus condition, the EEG was recorded as single epochs, filtered from either DC or 0.1 Hz to 200 Hz and sampled at a rate of 1.0 kHz.

Recording was initiated 100 ms prior to the acoustic onset and for 500 ms following the onset. Recording obtained for the VO stimuli used the same recording onset and offset as for the AV stimuli.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Off-line, the individual EEG single-sweeps were baseline corrected over the pre-stimulus interval and subjected to an automatic artifact rejection algorithm. A regression-based eye blink correction algorithm was applied to the accepted single sweeps (at least 1500 per subject), which were then averaged. The averages were filtered from 1 to 70 Hz and average-referenced. For each stimulus, data from all 12 subjects were used to generate grand average waveforms.

Source reconstruction

With Curry® software (Neuroscan, Texas), cortical activation reconstructions were generated by applying minimum norm least squares CDR analysis (Phillips et al., 1997). Source reconstruction utilized a three-shell, spherical head, volume conductor model with an outer radius of 9 cm. Analyses were constrained to the cortical surface of a segmented brain (Wagner et al., 1995). CDR was applied to the ERP data across the latency interval of 50-250 milliseconds; an interval that spanned both of the major activation peaks in the mean global field power for AV stimuli (see Figure 2). Results were computed at each millisecond of data. After filtering the averages from 1 to 70Hz, all information that is in the data is exposed, if results are presented in 1000/140 = 7msec steps, where 140 = 2 * 70Hz, that is, the Nyquist frequency. Results were computed in 1msec steps, because that was the sampling rate after filtering, and Curry™ 5 is not able to subsample. Solutions were derived using minimum norm least squares criteria in which the field is explained by a source configuration with minimum power. Depth weighting was applied, and regularization was used to achieve a fit error that matched the inverse of the signal-to-noise ratio (SNR) (Fuchs et al., 1999). Inverse SNRs for the five conditions were AO 0.0588, AVi 0.0455, AVc 0.0526, VO /bα/ 0.0526, and VO/gα/ 0.109. Obtained solutions were displayed with a common threshold across all conditions. The supplementary materials show animations for each condition comprising the CDRs at the millisecond rate at which they were computed. The cortical generators responsible for the signals of interest here likely were less dispersed than the resulting CDR activity. To partially account for this possibility, the results are described in terms of the foci of modeled highest current densities.

Results

The mean global field power for the AO, VO, AVc, and AVi responses shown in Figure 2 demonstrates that all the conditions produced structured activations, and that AO activity was earlier than VO activity. The overall level of activity was greater with the audiovisual (AVi and AVc) stimuli. Also the latencies of the peaks in the mean global field power varied across the different stimulus conditions.

Figure 3a-d shows the results in terms of the CDRs. Results are shown for the left and right hemispheres for AO /ba/, VO /ga/ and /ba/, AVi, and AVc. In Figure 3a and 3c, the CDRs are shown in 5-msec steps, beginning at 50 ms and continuing through to 100 ms; in Figure 3b and 3d, the CDRs are shown beginning in 20-msec steps from 100 to 240 ms, with a final time frame at 250 ms showing the return to subthreshold activation. The CDRs in the figure do not represent mean data across the intervals but single frames at the noted time points. All these data were thresholded to the same current density range across all conditions.

The foci here were stable results with continuity across time. None of the results discussed here relied on high-frequency potentials. The continuity of the events can be seen particularly well in the animations in the supplementary materials from 50 msec to 250 msec.

Auditory-only activity

Activation developed in the right temporal cortex at about 55-60 msec and continued through 180 msec (Figure 3c and d). The CDR analysis indicates that the peak in the MGFP at 165

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

msec for the AO condition is primarily attributable to activity in the right hemisphere (compare Figure 3b and 3d). Parietal activity that was more extensive on the right than the left was obtained in the interval 55 to 90 msec (Figure 3a and 3c).

AO activity centered on the left hemisphere superior temporal gyrus emerged in the vicinity of primary auditory cortex at 55 msec (Figure 3a). By 60-65 msec, activity had spread and included a small focus in the parietal lobule region as well as the inferior parietal lobule. Left inferior frontal activity that might correspond to the Broca's area pars triangularis (Amunts et al., 1999; Broca, 1861) developed at 65-70 msec and persisted for more than 30 ms before dissipating.

Left hemisphere activity appears to have resolved by 120 msec, while right hemisphere activity, particularly, in the inferior temporal region, persisted to 180 msec. The AO results demonstrate rapid development and resolution of processing (see also, Reale et al., 2007), as well as temporally early and simultaneous recruitment of parietal and frontal areas.

Visual-only activity

For the VO condition, only minor activation was present in the CDRs for either /bα/ or /gα/ until approximately 120 msec (Figures 3b and 3d). However, in the MGFP analyses, low-level occipital activity was present beginning at 65 msec and diminishing at 80 msec, particularly for visual /bα/ (Figure 2). This is barely noticeable as a small area of activation in the CDR plots for visual /bα/ at 70 and 75 ms (Figure 3a and 3c).

The much more extensive later occipital activation corresponds to the increased variance displayed in the MGFP curves in Figure 2. After 120 msec there was a spread of activation from the occipital pole for both VO /gα/ and /bα/ stimuli. In the left hemisphere, this pattern of activation is quite similar across both VO stimuli. On the right, CDR activity for both stimuli also originated at the occipital pole, but expanded much more broadly across the lateral surface of posterior cortex for VO /bα/ stimulation. Activity also spread extensively in right inferior parietal and inferior temporal cortex, particularly, at the latencies of 160 through 200 msec.

The VO results demonstrate the temporally extended development and resolution of activation, as well as recruitment of some areas beyond the occipital cortex but not into auditory temporal cortex (Reale et al., 2007). Although some low-level activity appears at early latencies, the more spatially extensive and higher current densities occurred at later latencies.

AVc left hemisphere activity

In the left hemisphere (Figure 3a), at 55 msec, activity in the region of the auditory cortex appears to have been suppressed (relative to the AO condition), and two small foci of activity appeared, one modeled in posterior parietal cortex, an area consistent with the IPS, and the other in the supramarginal gyrus (SMG). By 60-65 msec, these areas had expanded. The activity then also included the STG and the angular gyrus (AG). Across 60 to 80 msec, AVc temporal and parietal cortical activation was similar to AO activation patterns. In the interval of 70-95 msec, low amplitude inferior frontal (IF) cortex activity was present. At 85 msec, activity in dorsolateral prefrontal cortex (PFC) can be seen, and it expanded in area and amplitude and continued through 100 msec. Occipital activity was recorded beginning at approximately 90 ms and persisted throughout the time epoch under examination. This activity appears to be delayed relative to the activity in the CDRs for the VO /bα/ noted above. Low amplitude inferior temporal gyrus (ITG) activity was recorded in the interval 100-120 msec.

Relatively late, 120-160 msec, activity extended from the occipital cortex anteriorly to the posterior STS. Evidence for focused STS activity was obtained in the interval of 120-180 msec.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

However, the current density magnitudes were not large. From 160 to 200 msec, unlike either the AO or VO conditions, activity was focused in the region of the SMG and AG.

AVi left hemisphere activity

Across the period of 50 to 70 msec, the left hemisphere AVi activity was dissimilar from the AO activity, even though the acoustic stimulus was the same in the two conditions (Figure 3a). Early activity in the region of the auditory cortex appears to have been suppressed beyond the period noted for the AVc stimulus. At 50 msec, very low amplitude activity was present in dorsolateral prefrontal cortex and the region of the middle STS, and the occipital cortex. Dorsolateral PFC activity continued to be recorded through 60 msec. But middle STS activity was not visible at 55 msec. More persistent and more extensive activity developed first in ITG cortex, seen at 65 msec and then was present in IF cortex at 75 msec. Superior parietal activation emerged at 75 msec and persisted through 120 msec. This activity is consistent with an IPS source. Between 65 msec and 100 msec, focused activity moved from the inferior temporal lobe to the parietal lobe in the region of the angular and supramarginal gyri. Weak activity in dorsolateral prefrontal cortex arose again at 85 msec and persisted to 100 msec. Similar to the results for the AVc stimulus, activity was modeled in parietal cortex, consistent with SMG and AG between approximately 160 and 220 msec, with weak activation seen at 140 msec. The SMG/AG activity could be viewed as a continuation of the earlier activity in this area.

AVc and AVi right hemisphere activity

Generally, there were many similarities between the right hemisphere AVi and AVc activations. However, the former appeared somewhat delayed in terms of the extent of spread of activity and also in terms of the onset of dorsolateral PFC activity. Also, initial AVc activity was generally similar to AO activity from 60 through 100 msec. But initial AVi activity was generally different from that of AO activity, with an initial parietal focus (50 msec) and early persistent occipital activity (55 to 70 msec).

Specifically, looking in more detail, at 50 msec, for the AVi stimulus, a brief focus in the superior parietal area was observed. At 60 msec, for the AVc stimulus, dorsolateral PFC activity emerged, moved briefly inferiorly to the inferior frontal gyrus, and returned to the dorsolateral PFC were it was sustained through 100 msec. These frontal activations differed from the AVi condition, for which the earliest (75 msec) focused frontal activity was in IF gyrus and was sustained through 120 msec.

At approximately 75 msec, the activation patterns were qualitatively similar across the AO, AVi, and AVc conditions and remained so through 95-100 msec, except for the more superior, widespread dorsolateral PFC activations of AVc. Activity in the right STS extended temporally for the AO, AVc, and AVi stimuli. At 65 msec, with the AVc stimulus, there was more posterior STS activation, and more extensive activity continued to approximately to 120 or 140 msec. The AVi stimulus resulted in a later onset of STS activity (at approximately 80 msec).

In the AVi condition, occipital activity was modeled from 55 to 70 msec and then from 85 to 240 msec. In the AVc stimulus, occipital activity was modeled from 60 msec through 240 msec, however, this activity was not consistently present at the occipital pole. At 120 msec, and continuing through the remaining time epoch, AVc and AVI activity appeared as a composite of the activated areas for the AO and the VO /bα/ models. In contrast with the left hemisphere, where an emergent parietal activation was obtained between 160 and 220 msec, the right hemisphere activity was merely more extensive temporally and spatially in the AVc and AVi conditions than the AO and VO conditions.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Summary: AVc and AVi

Summarizing the results for the left hemisphere (Figure 4a): Differences and similarities were seen across the AVi and AVc stimulus conditions. AVi activations were temporally later by as many as 20 msec in the SMG/AG and the IPS. The temporal differences in STG, STS, and IF cortex across AVi and AVc were smaller and closer to the resolution of the measurements. Patterns of early latency frontal activity varied across audiovisual conditions. Dorsolateral PFC activity was earlier for AVi than AVc.

Evidence for the temporal aspect of the Miller and d'Esposito (2005) hypothesis was weak. The STS was not the focus of earliest activity, nor was it prominent at later latencies. The left dorsolateral PFC, the IPS, and the SMG/AG areas became active earlier and at higher amplitude than the left STS, contrary to prediction. The late latency left activation patterns departed most dramatically from the AO and VO conditions, with the addition of a focal activation of the SMG/AG area from 160 to 220 msec.

In the right hemisphere (Figure 4b), AVi activations generally had later onsets than AVc activations. Strikingly, many of the focal areas became active in close temporal correspondence in the interval from 70-80 msec (SMG/AG, IPS, STG, ITG, STS, IF), again contrary to the prediction. Other than the difference across conditions in latency, notably, the frontal activity varied, with the dorsolateral PFC focus for the AVc condition and the more inferior focus for the AVi condition.

Discussion

EEG generally affords better temporal (7-msec here) than spatial (1-2-cm here) resolution in comparison with, for example, fMRI. Given that *caveat*, it can be said that several striking features of spatio-temporal multisensory cortical activation were identified. First, left STS was not prominent in the spatio-temporal dynamics of audiovisual stimulus processing. Although STS activation was obtained with AV stimuli, it was generally not the earliest, highest current density, nor most spatially focal. Right STS showed activation that was similar across AO, AVc, and AVi stimuli. Second, evidence for frontal and parietal activity during AV conditions was obtained at the earliest latencies examined. Third, activity in latencies from as early as approximately 65 to 100 msec was concurrently present in frontal, parietal and temporal cortex. Fourth, persistent activity, unique to the AV conditions, emerged at approximately 160 msec and was sustained for tens of milliseconds in a left hemisphere area consistent with SMG/AG. Fifth, the spatio-temporal dynamics of AO and VO speech stimulus processing differed, the former mostly distinct from the latter, and developing and resolving much more quickly.

Hypothesized spatio-temporal dynamics

The large-scale regional and temporal patterns of activity reported here were used to address the hypothesis posed by Miller and d'Esposito (2005) that audiovisual speech integration is initiated in the middle STS, with recruitment of intraparietal areas when stimuli are not in their normal correspondence, and with additional processing efforts resulting in recruitment of Broca's area. While activation was obtained within the hypothesized areas, its pattern was not consistent with the hypothesized temporal dimension. Complex temporal patterns were obtained, including concurrent activations. An implication of the results here is that the functional attributions of the hypothesized circuit, tied as they were to initial integration followed by additional processing dependent on stimulus difficulty, appear not to be instantiated in the spatio-temporal dynamics of audiovisual stimuli similar to those in Miller and d'Esposito (2005).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Several previous studies have reported STS to be an audiovisual integration site (Callan et al., 2003; Calvert et al., 2000; Miller and d'Esposito, 2005; Sekiyama et al., 2003; Wright et al., 2003). But others have noted the absence of differential STS activation. For example, matched and mismatched vowels failed to result in differential STS activation (Ojanen et al., 2005) (see also, Jones and Callan, 2003; Olson et al., 2002). Here we obtained a broad area of left activation from AG/SMG to the MTG, centered more superiorly, at approximately 160 to 180 msec in both AV conditions. While STS appears to participate in audiovisual integration, our results do not suggest it to be a major orchestrator of cortical activations.

Audiovisual integrative processing

Peaks of activity in the left posterior parietal cortex, labeled “IPS” here, might correspond to the intraparietal activation in Miller and d'Esposito (2005). Evidence exists for forward connections to the dorsolateral PFC from multisensory IPS (Pandya and Yeterian, 1996; Sugihara et al., 2006). However, there was early latency evidence for this pathway in the AO condition in addition to the two AV conditions, suggesting that this connection is not serving a uniquely cross-modal integrative function.

The longer latency (particularly, 160-220 msec) activations centered in left SMG/AG were not predicted by Miller and d'Esposito (2005), nor by the VO and AO activation patterns. However, previous evidence suggests that this site is concerned with multisensory integration (Jones and Callan, 2003; Kaiser et al., 2006). This area of association cortex (which appears to include Wernicke's area) has been hypothesized to function as a transmodal gateway that binds or associates bottom-up modality-specific, representational patterns during perception (Bernstein et al., 2004a; Mesulam, 1998). The persistence and later emergence of activity in SMG/AG, as well as its stimulus-related differences across AVi and AVc support the suggestion that the response in this region is critical in audiovisual speech integration.

Implications for the spatio-temporal dynamics of audiovisual phonetic integration

The differences reported here between the VO and AO activations, with the former mostly in posterior cortex and the latter mostly in temporal cortex, suggest that there is not a single phonetic speech information processing area. However, considerable controversy has surrounded whether bottom-up visible speech features are processed by the auditory cortical areas that process audible speech (Bernstein et al., 2002; Calvert et al., 1997; Campbell et al., 2001; Paulesu et al., 2003; Pekkola et al., 2005).

The passive task here might be responsible for the relatively restricted activity in posterior areas with VO speech stimuli. With an active speech perception task, activations resulting from VO stimuli might have extended into left anterior temporal regions typically associated with auditory speech processing (Scott and Johnsrude, 2003). However, recently, a study on posterolateral STG using intra-cortical recordings showed little response to VO speech stimulation when the subject was performing a monitoring task (Reale et al., 2007). In the same study, audiovisual stimuli were compared for which the visual stimulus was either a nonsense syllable or a face motion. The results showed little evidence for sensitivity to the visual dimension in the posterolateral STS region, consistent with the results here.

The ERP literature on audiovisual processing of simple non-speech stimuli has noted early exogenous (prior to 100 msec) interactions (Giard and Peronnet, 1999; Molholm et al., 2002). But in accounting for audiovisual speech stimulus processing, temporal constraints related to stimulus complexity should be taken into account. Time is needed not only to process information but also for the transduction of the information to the relevant processing locations. In general, stimuli comprising multiple stimulus attributes are thought to be processed hierarchically following lower-level feature analyses.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Previous studies have established that the first volley of stimulus-driven activity into the auditory core cortex occurs around 11-20 msec post stimulus onset (Saron et al., 2001; Steinschneider et al., 1999; Yvert et al., 2001; Yvert et al., 2005). Based on intra-cerebral recordings, simple auditory stimuli, and similar source localization approaches to those reported here, several simultaneous distinct sites of activation have been observed, including, Heschl's gyrus, the planum temporale, and STG, all active within 28 to 100 msec post-stimulus onset (Yvert et al., 2005). The site of auditory phonetic processing remains controversial, but researchers seem to agree that it is not the areas that are activated temporally earliest in the bottom-up synaptic hierarchy of the auditory cortex (Hickok and Poeppel, 2007; Scott and Johnsrude, 2003). It has been estimated that the conscious auditory speech percept develops within 150-200 msec post-stimulus onset (Näätänen, 2001), consistent with a locus distal from the primary auditory area.

Intra-cortical recordings in V1/V2 have shown the earliest stimulus-driven response to be at approximately 45-60 msec (Foxe and Simpson, 2002; Krolak-Salmon et al., 2001). Trans-cortical feature processing requires time (Schroeder and Foxe, 2002). Latency for combining visual form and motion at the level of the cortex is at least 100 msec, and face motion processing might require closer to 170 msec (Puce and Perrett, 2003).

Thus, latency measures suggest that audiovisual speech stimulus processing, which likely involves the hierarchically processed extraction and integration of phonetic features, could require *at least* 100-150 msec from stimulus onset of both auditory and visual speech stimuli. Furthermore, evidence in the literature that seems unambiguously relevant to audiovisual phonetic processing, for which incongruent audiovisual speech stimuli were presented, has implicated relatively long latencies for speech integration, although non-speech controls were not tested. For example, when participants were asked to detect audiovisual vowel incongruity, a negative response around 300 msec was obtained, which was attributed to the difficulty of integrating conflicting speech information (Lebib et al., 2004). In another study, when congruent and incongruent audiovisual vowels were presented, early (85 msec) responses were evidence for audiovisual interactions independent of the vowel identities, but differential processing related to vowel congruity was observed at 155 msec (Klucharev et al., 2003). These longer latencies are commensurate here with the latencies of the left hemisphere emergent activity in SMG/AG (see Figure 3b and 3d) and persisting between 140 and 220 msec.

Early frontal activity

With audiovisual stimuli, areas of dorsolateral PFC and inferior frontal cortex activity in both hemispheres arose early. Frontal activity was, however, very similar to that in the AO condition, raising the possibility that activation during AV conditions was attributable to the auditory stimulus. The P50 auditory evoked potential (also known as P1, Pb, or Pb1) is obtained in the interval 45-75 msec post-stimulus onset. The P50 can be reduced when a train of auditory stimuli is presented, and this effect has been interpreted as sensory gating that screens out redundant information, thus, theoretically freeing capacity for higher-level processing (Korzyukov et al., 2007; Lebib et al., 2004). The three conditions here with audio have a repeated syllable, likely to elicit sensory gating. Recent intracortical recordings on human epilepsy patients have shown a frontal generator for auditory sensory gating associated with the P50 response (Korzyukov et al., 2007). Thus, the early frontal activity here could be attributable to the repetitive auditory stimulus rather than to an early integrative process (Giard and Peronnet, 1999; Lebib et al., 2004). Interpretation of early frontal effects being due to stimulus repetition or other global factors such as anticipation (Teder-Salejarvi et al., 2002) is consistent with our hypothesis that audiovisual phonetic integration follows at longer latencies.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Limitations of the CDRs

A valid question here is the extent to which the CDR spatial resolution is adequate for examining spatiotemporal circuit hypotheses for audiovisual speech processing. The intrinsic under-determination in the CDRs due to the inverse problem leads to the possibility of regional spatial errors. In addition, application of larger numbers of electrodes could have afforded greater spatial resolution. One piece of converging evidence in favor of accepting the regional results here is the extent to which they appear accurate for the AO and VO stimuli.

Interestingly, some previous studies have shown little structured visual-only ERPs (Besle et al., 2004; van Wassenhove et al., 2005). Here there were extensive structured ERPs and relatively early latency activation to the visual speech stimuli (see also, Saint-Amour et al., 2007). In the present study, a large number of EEG sweeps was collected, increasing the signal-to-noise ratios and affording good estimates of the ERPs. Given the substantial results in the EEG literature on non-speech face motion (Puce et al., 2007; Puce et al., 2000), a reasonable expectation is that we would obtain visual cortex activity in CDRs with speech face motion, as was the case.

Nevertheless, the underlying cortical generators responsible for the observed activity were likely less spatially dispersed than those visualized in the CDRs. Taking this into account, the results were described in terms of the anatomical locations with high current densities. An alternative approach would have been to undertake analyses such as the permutation methods presented by Pantazis et al. (2005), which could possibly reduce the spatial extent of activity by statistical thresholding, although the specific extent of blurring is intrinsic to the inverse operator. The Pantazis-et-al. method requires equivalent duration pre- and post-stimulus data, which were not available here but could be obtained in the future.

Alternatively, our interest in the hypothesis of Miller and d'Esposito (Miller and d'Esposito, 2005) calls for an approach that is arguably less conservative (relatively lenient) regarding Type I error. That is, we sought evidence for the possibility that the hypothesized circuit was true. Our result supported the activation of the hypothesized regions but not their hypothesized temporal dynamics.

Measures of audiovisual integration

Super-additivity, for which the sum of activity recorded from single neurons responding to unisensory stimulation is less than the activity to multisensory stimulation, has been until recently regarded as, perhaps, *the* signature of multisensory integration (Meredith, 2002; Stein and Meredith, 1993). Similar computations using the BOLD signals of fMRI or using ERPs with surface electrodes are problematic. These measures reflect large populations of neurons, and the proportion of multisensory neurons in cortex is estimated to be relatively small (Laurienti et al., 2005). Quantitative measures obtained under multisensory conditions could appear to demonstrate super-additivity attributable to integration when in fact two or more populations of sensory-specific neurons were activated in a particular location (Beauchamp et al., 2004; Calvert, 2001). Different computations of multisensory integration, including super-additivity can result in qualitatively different activation patterns (Beauchamp, 2005).

In some electrophysiological studies, audiovisual integration has been evaluated using measures of super-additivity or response suppression that involved adding waveforms from different unisensory stimulation conditions and comparing them to waveforms from the multisensory stimulation condition (Klucharev et al., 2003; van Wassenhove et al., 2005). This method is open to the possibility that activity on individual electrode recordings reflects additivity across near and far field electrical sources (Besle et al., 2004). It also risks measurement bias, in that audiovisual processing tends to be offset temporally from unisensory

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

processing (Giard and Peronnet, 1999; Molholm et al., 2002), as noted here, such that the peaks of activity across conditions do not coincide: Multisensory super-additivity could arise, for example, because the comparison among conditions takes the peak of audiovisual processing and compares it with activity that peaks earlier or later in time. The bias in this case is towards confirmation of super-additivity.

More recently, comparisons have been made directly across unisensory versus bisensory conditions (Reale et al., 2007; van Atteveldt et al., 2007; van Wassenhove et al., 2005). The study reported here contributes to the literature on audiovisual processing that does not depend on addition and subtraction to obtain sub- or super-additivity. Here CDR models across conditions were compared directly, without recourse to subtraction or addition of waveforms or of modeled parameters. This approach revealed qualitatively different spatio-temporal patterns, including emergent activity not predicted by activity in AO or VO conditions.

Conclusion

Evidence was sought for the circuit supporting audiovisual speech integrative in the comparison across AO, VO, and AV conditions expressed as CDR models. Qualitatively different spatiotemporal activation patterns across different audiovisual conditions, and between audiovisual and unisensory conditions were observed. The use of a temporally sensitive measure, current density reconstructions, demonstrated a dynamically distributed, including simultaneous, pattern of activations involving areas that had been hypothesized to be activated sequentially (Miller and d'Esposito, 2005). Early latency dorsolateral prefrontal and inferior frontal cortex as well as SMG/AG and IPS activations were obtained with AV and AO speech. Activation in SMG/AG that was not predicted in AO and VO conditions was observed again at relatively long latencies and is hypothesized here to be related to the integration of the auditory and visual speech stimulus information.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by the NSF (IIS9996088) and the NIH (DC006035) (Bernstein, PI).

References

Amunts K, Schleicher A, Burgel U, Mohlberg H, Uylings HB, Zilles K. Broca's region revisited: cytoarchitecture and intersubject variability. Journal of Comparative Neurology 1999;412:319–341. [PubMed: 10441759]

Arnold P, Hill F. Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. British Journal of Psychology 2001;92:339–355.

Beauchamp MS. Statistical criteria in FMRI studies of multisensory integration. Neuroinformatics 2005;3:93–113. [PubMed: 15988040]

Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A. Unraveling multisensory integration: patchy organization within human STS multisensory cortex. Nature Neuroscience 2004;7:1190–1192.

Bernstein, LE.; Auer, ET., Jr.; Moore, JK. Audiovisual Speech Binding: Convergence or Association?. In: Calvert, GA.; Spence, C.; Stein, BE., editors. Handbook of Multisensory Processing. MIT; Cambridge, MA: 2004a. p. 203-223.

Bernstein LE, Auer ET Jr. Moore JK, Ponton CW, Don M, Singh M. Visual speech perception without primary auditory cortex activation. NeuroReport 2002;13:311–315. [PubMed: 11930129]

Bernstein LE, Takayanagi S, Auer ET Jr. Auditory speech detection in noise enhanced by lipreading. Speech Communication 2004b;44:5–18.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Besle J, Fort A, Delpuech C, Giard M-H. Bimodal speech: early suppressive visual effects in human auditory cortex. European Journal of Neuroscience 2004;20:2225–2234. [PubMed: 15450102]

Breeuwer M, Plomp R. Speechreading supplemented with auditorily presented speech parameters. Journal of the Acoustical Society of America 1986;79:481–499. [PubMed: 3950202]

Broca P. Sur le siége de la faculté du langage articulé, suivies d'une observation d'aphemie (perte de la parole). Bulletins de la Société d'Anatomie (Paris) 1861;6:330–357.

Buckner RL. The hemodynamic inverse problem: Making inferences about neural activity from measured MRI signals. Proceedings of the National Academy of Sciences of the United States of America 2003;100:2177–2179. [PubMed: 12606715]

Callan DE, Jones JA, Munhall K, Callan AM, Kroos C, Vatikiotis Bateson E. Neural processes underlying perceptual enhancement by visual speech gestures. NeuroReport 2003;14:2213–2218. [PubMed: 14625450]

Calvert GA. Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cerebral Cortex 2001;11:1110–1123. [PubMed: 11709482]

Calvert GA, Brammer MJ, Bullmore ET, Campbell R, Iversen SD, David AS. Response amplification in sensory-specific cortices during crossmodal binding. NeuroReport 1999;10:2619–2623. [PubMed: 10574380]

Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire PK, Woodruff PW, Iversen SD, David AS. Activation of auditory cortex during silent lipreading. Science 1997;276:593–596. [PubMed: 9110978]

Calvert GA, Campbell R, Brammer MJ. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Current Biology 2000;10:649–657. [PubMed: 10837246]

Campbell R, MacSweeney M, Surguladze S, Calvert G, McGuire P, Suckling J, Brammer MJ, David AS. 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). Cognitive Brain Research 2001;12:233–243. [PubMed: 11587893]

Colin C, Radeau M, Soquet A, Deltenre P. Generalization of the generation of an MMN by illusory McGurk percepts: voiceless consonants. Clinical Neurophysiology 2004;115:1989–2000. [PubMed: 15294201]

Colin C, Radeau M, Soquet A, Demolin D, Colin F, Deltenre P. Mismatch negativity evoked by the McGurk-MacDonald effect: a phonetic representation within short-term memory. Clinical Neurophysiology 2002;113:495–506. [PubMed: 11955994]

Erber NP. Auditory-visual perception of speech. Journal of Speech and Hearing Disorders 1975;40:481–492. [PubMed: 1234963]

Foxe JJ, Simpson GV. Flow of activation from V1 to frontal cortex in humans. A framework for defining “early” visual processing. Experimental Brain Research 2002;142:139–150.

Fuchs M, Wagner M, Kohler T, Wischmann H-A. Linear and nonlinear current density reconstruction. Journal of Clinical Neurophysiology 1999;16:267–295. [PubMed: 10426408]

Giard MH, Peronnet F. Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. Journal of Cognitive Neuroscience 1999;11:473–490. [PubMed: 10511637]

Grant KW. The effect of speechreading on masked detection thresholds for filtered speech. Journal of the Acoustical Society of America 2001;109:2272–2275. [PubMed: 11386581]

Grant KW, Seitz PF. The use of visible speech cues for improving auditory detection of spoken sentences. Journal of the Acoustical Society of America 2000;108:1197–1208. [PubMed: 11008820]

Grant KW, Walden BE. Evaluating the articulation index for auditory-visual consonant recognition. Journal of the Acoustical Society of America 1996;100:2415–2424. [PubMed: 8865647]

Grant KW, Walden BE, Seitz PF. Auditory-visual speech recognition by hearing-impaired subjects: consonant recognition, sentence recognition, and auditory-visual integration. Journal of the Acoustical Society of America 1998;103:2677–2690. [PubMed: 9604361]

Hickok G, Poeppel D. The cortical organization of speech processing. Nature Reviews: Neuroscience 2007;8:393–402.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Iverson P, Bernstein LE, Auer ET Jr. Modeling the interaction of phonemic intelligibility and lexical structure in audiovisual word recognition. Speech Communication 1998;26:45–63.

Jasper HH. The ten-twenty electrode system of the International Federation. Electroencephalography and Clinical Neurophysiology 1958;10:371–375.

Jones JA, Callan DE. Brain activity during audiovisual speech perception: An fMRI study of the McGurk effect. NeuroReport 2003;14:1129–1133. [PubMed: 12821795]

Kaiser J, Hertrich I, Ackermann H, Lutzenberger W. Gamma-band activity over early sensory areas predicts detection of changes in audiovisual speech stimuli. Neuroimage 2006;30:1376–1382. [PubMed: 16364660]

Kishon-Rabin L, Boothroyd A, Hanin L. Speechreading enhancement: A comparison of spatial-tactile display of voice fundamental frequency (F0) with auditory F0. Journal of the Acoustical Society of America 1996;100:593–602. [PubMed: 8675850]

Klucharev V, Mottonen R, Sams M. Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. Cognitive Brain Research 2003;18:65–75. [PubMed: 14659498]

Korzyukov O, Pflieger ME, Wagner M, Bowyer SM, Rosburg T, Sundaresan K, Elger CE, Boutros NN. Generators of the intracranial P50 response in auditory sensory gating. Neuroimage 2007;35:814–826. [PubMed: 17293126]

Krolak-Salmon P, Henaff MA, Tallon-Baudry C, Yvert B, Fischer C, Vighetto A, Betrand O, Mauguiere F. How fast can the human lateral geniculate nucleus and visual striate cortex see? Society for Neuroscience Abstracts 2001;27:913.

Laurienti PJ, Perrault TJ, Stanford TR, Wallace MT, Stein BE. On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. Experimental Brain Research 2005;166:289–297.

Lebib R, Papo D, Douiri A, de Bode S, Dowens MG, Baudonniere P-M. Modulations of ‘late’ event-related brain potentials in humans by dynamic audiovisual speech stimuli. Neuroscience Letters 2004;372:74–79. [PubMed: 15531091]

MacLeod A, Summerfield Q. Quantifying the contribution of vision to speech perception in noise. British Journal of Audiology 1987;21:131–141. [PubMed: 3594015]

McGurk H, MacDonald J. Hearing lips and seeing voices. Nature 1976;264:746–748. [PubMed: 1012311]

Meredith MA. On the neuronal basis for multisensory convergence: a brief overview. Cognitive Brain Research 2002;14:31–40. [PubMed: 12063128]

Mesulam MM. From sensation to cognition. Brain 1998;121:1013–1052. [PubMed: 9648540]

Miller LM, d'Esposito M. Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. Journal of Neuroscience 2005;25:5884–5893. [PubMed: 15976077]

Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ. Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. Cognitive Brain Research 2002;14:115–128. [PubMed: 12063135]

Moody-Antonio S, Takayanagi S, Masuda A, Auer J,ET, Fisher L, Bernstein LE. Improved speech perception in adult prelingually deafened cochlear implant recipients. Otology and Neurotology 2005;26:649–654. [PubMed: 16015162]

Näätänen R. The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). Psychophysiology 2001;38:1–21. [PubMed: 11321610]

Ojanen V, Mottonen R, Pekkola J, Jaaskelainen IP, Joensuu R, Autti T, Sams M. Processing of audiovisual speech in Broca's area. Neuroimage 2005;25:333–338. [PubMed: 15784412]

Olson IR, Gatenby JC, Gore JC. A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. Cognitive Brain Research 2002;14:129–138. [PubMed: 12063136]

Pandya DN, Yeterian EH. Comparison of prefrontal architecture and connections. Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences 1996;351:1423–1432.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Pantazis D, Nichols TE, Baillet S, Leahy RM. A comparison of random field theory and permutation methods for the statistical analysis of MEG data. Neuroimage 2005;25:383–394. [PubMed: 15784416]

Paulesu E, Perani D, Blasi V, Silani G, Borghese NA, De Giovanni U, Sensolo S, Fazio F. A functional-anatomical model for lipreading. Journal of Neurophysiology 2003;90:2005–2013. [PubMed: 12750414]

Pekkola J, Ojanen V, Autti T, Jaaskelainen IP, Mottonen R, Tarkianen A, Sams M. Primary auditory cortex activation by visual speech: an fMRI study at 3T. NeuroReport 2005;16:125–128. [PubMed: 15671860]

Phillips JW, Leahy RM, Mosher JC. MEG-based imaging of focal neuronal current sources. IEEE Transactions on Medical Imaging 1997;16:338–348. [PubMed: 9184896]

Puce A, Epling JA, Thompson JC, Carrick OK. Neural responses elicited to face motion and vocalization pairings. Neuropsychologia 2007;45:93–106. [PubMed: 16766000]

Puce A, Perrett D. Electrophysiology and brain imaging of biological motion. Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences 2003;358:435–445.

Puce A, Smith A, Allison T. ERPs evoked by viewing facial movements. Cognitive Neuropsychology 2000;17:221–239.

Reale RA, Calvert GA, Thesen T, Jenison RL, Kawasaki H, Oya H, Howard MA, Brugge JF. Auditory-visual processing represented in the human superior temporal gyrus. Neuroscience 2007;145:162–184. [PubMed: 17241747]

Reisberg, D.; McLean, J.; Goldfield, A. Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In: Dodd, B.; Campbell, R., editors. Hearing by Eye: The Psychology of Lip-reading. Lawrence Erlbaum; London: 1987. p. 97-113.

Saint-Amour D, De Sanctis P, Molholm S, Ritter W, Foxe JJ. Seeing voices: High-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion. Neuropsychologia 2007;45:587–597. [PubMed: 16757004]

Saito DN, Yoshimura K, Kochiyama T, Okada T, Honda M, Sadato N. Cross-modal binding and activated attentional networks during audio-visual speech integration: A functional MRI study. Cerebral Cortex 2005;15:1750–1760. [PubMed: 15716468]

Saron CD, Molholm S, Ritter W, Murray MM, Schroeder CE, Foxe JJ. Possible auditory activation of visual cortex in a simple reaction time task: A high density ERP study. Society for Neuroscience Abstracts 2001;27:1795.

Schroeder CE, Foxe JJ. The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. Cognitive Brain Research 2002;14:187–198. [PubMed: 12063142]

Scott SK, Johnsrude IS. The neuroanatomical and functional organization of speech perception. Trends in Neurosciences 2003;26:100–107. [PubMed: 12536133]

Sekiyama K, Sugita Y, Kanno I, Miura S. Auditory-visual speech perception examined by functional MRI and PET. Neuroscience Research 2003;47:277–287. [PubMed: 14568109]

Stein, BE.; Meredith, MA. The Merging of the Senses. MIT; Cambridge, MA: 1993.

Steinschneider M, Volkov IO, Noh MD, Garell PC, Howard MA 3rd. Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. Journal of Neurophysiology 1999;82:2346–2357. [PubMed: 10561410]

Sugihara T, Diltz MD, Averbeck BB, Romanski LM. Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. Journal of Neuroscience 2006;26:11138–11147. [PubMed: 17065454]

Sumby WH, Pollack I. Visual contribution to speech intelligibility in noise. Journal of the Acoustical Society of America 1954;26:212–215.

Teder-Salejarvi WA, McDonald JJ, Di Russo F, Hillyard SA. An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. Cognitive Brain Research 2002;14:106–114. [PubMed: 12063134]

van Atteveldt NM, Formisano E, Goebel R, Blomert L. Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex. Neuroimage 2007;36:1345–1360. [PubMed: 17513133]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

van Wassenhove V, Grant KW, Poeppel D. Visual speech speeds up the neural processing of auditory speech. Proceedings of the National Academy of Sciences of the United States of America 2005;102:1181–1186. [PubMed: 15647358]

Wagner, M.; Fuchs, M.; Wischmann, H-A.; Ottenberg, K.; Dössel, O. Cortex segmentation from 3D MR images for MEG reconstructions. In: Baumgartner, C.; Deecke, L.; Stroink, G.; Williamson, SJ., editors. Biomagnetism: Fundamental Research and Clinical Applications. Elsevier Science IOS Press; Amsterdam, The Netherlands: 1995. p. 433-438.

Wright TM, Pelphrey KA, Allison T, McKeowin MJ, McCarthy G. Polysensory interactions along lateral temporal regions evoked by audiovisual speech. Cerebral Cortex 2003;13:1034–1043. [PubMed: 12967920]

Yvert B, Crouzeix A, Bertrand O, Seither-Preisler A, Pantev C. Multiple supratemporal sources of magnetic and electric auditory evoked middle latency components in humans. Cerebral Cortex 2001;11:411–423. [PubMed: 11313293]

Yvert B, Fischer C, Bertrand O, Pernier J. Localization of human supratemporal auditory areas from intracerebral auditory evoked potentials using distributed source models. Neuroimage 2005;28:140–153. [PubMed: 16039144]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

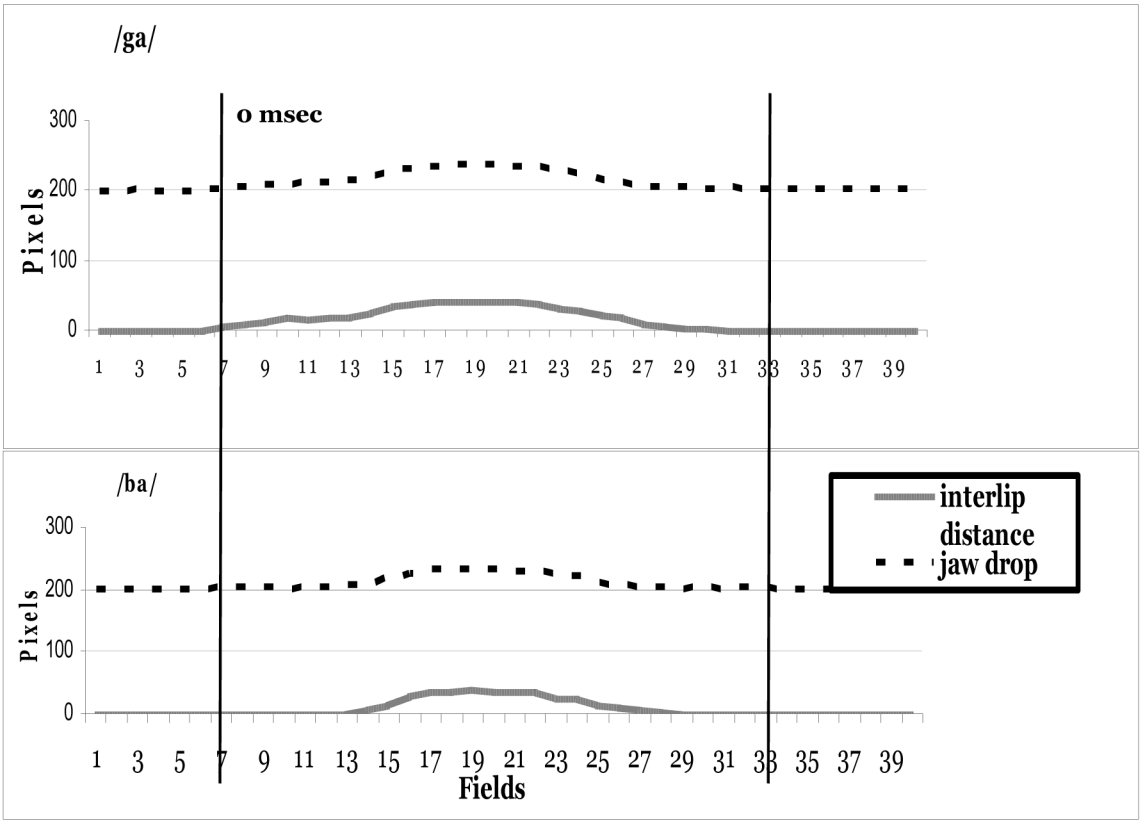


Figure 1.
Interlip distance and jaw drop in pixels. The vertical 0-msec line corresponds to the onset of the acoustic signal and the 0-msec EEG sampling point.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

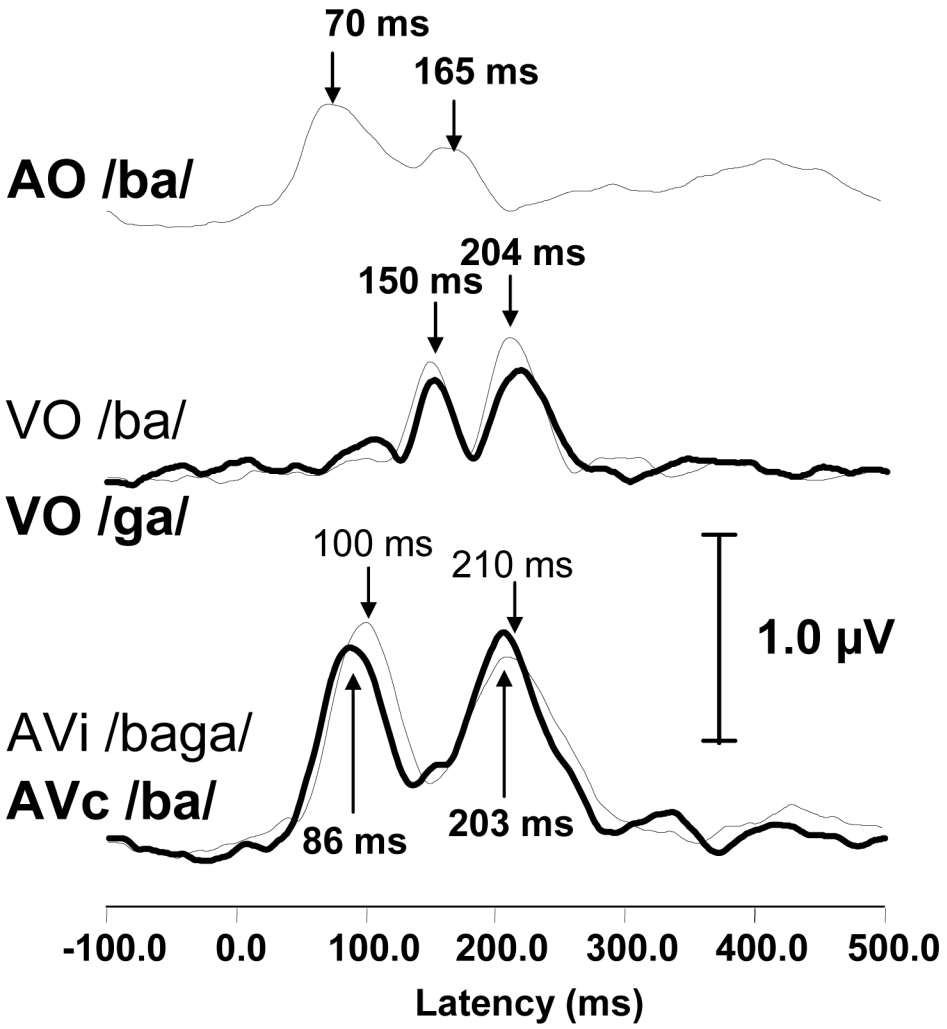


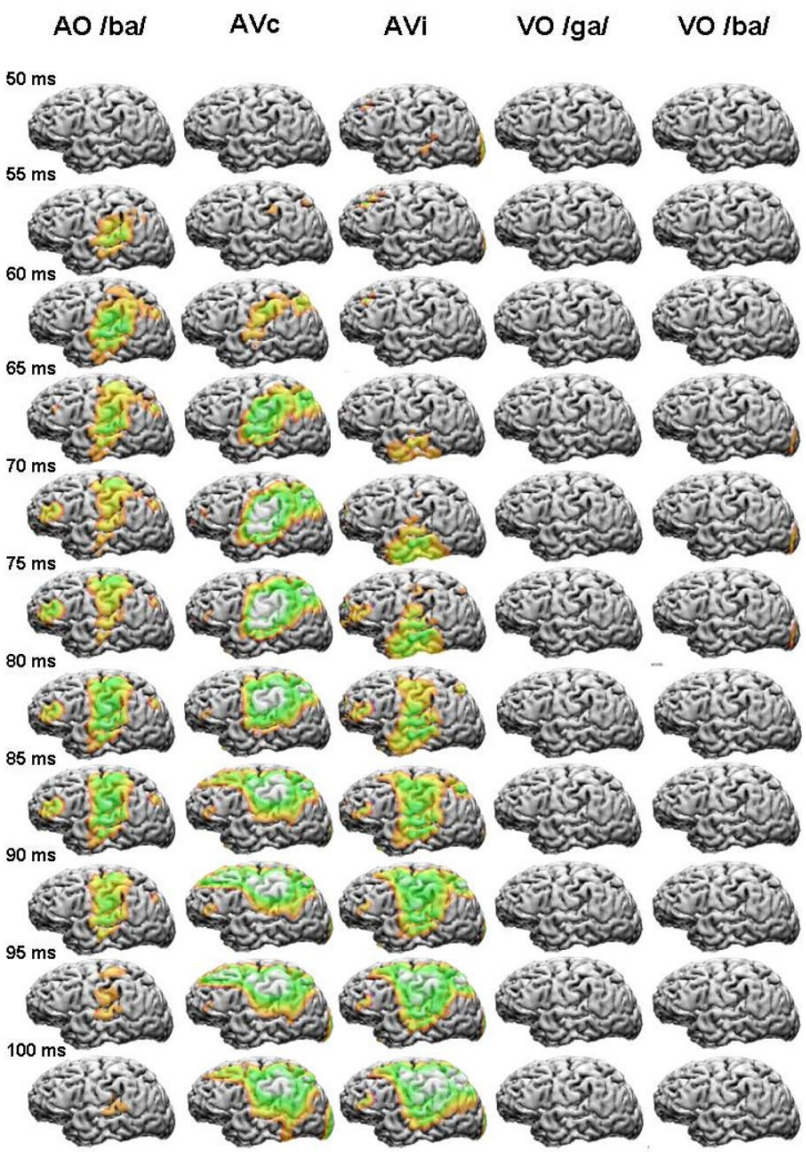
Figure 2.
Mean global field power (MGFP) for the grand mean ERPs in the AO, VO, and the two AV conditions of congruent AV stimulation (auditory /ba/ and visual /ba/) and incongruent AV stimulation (auditory /ba/ and visual /ga/). Each stimulus condition produced two distinct peaks in the MGFP.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

3a

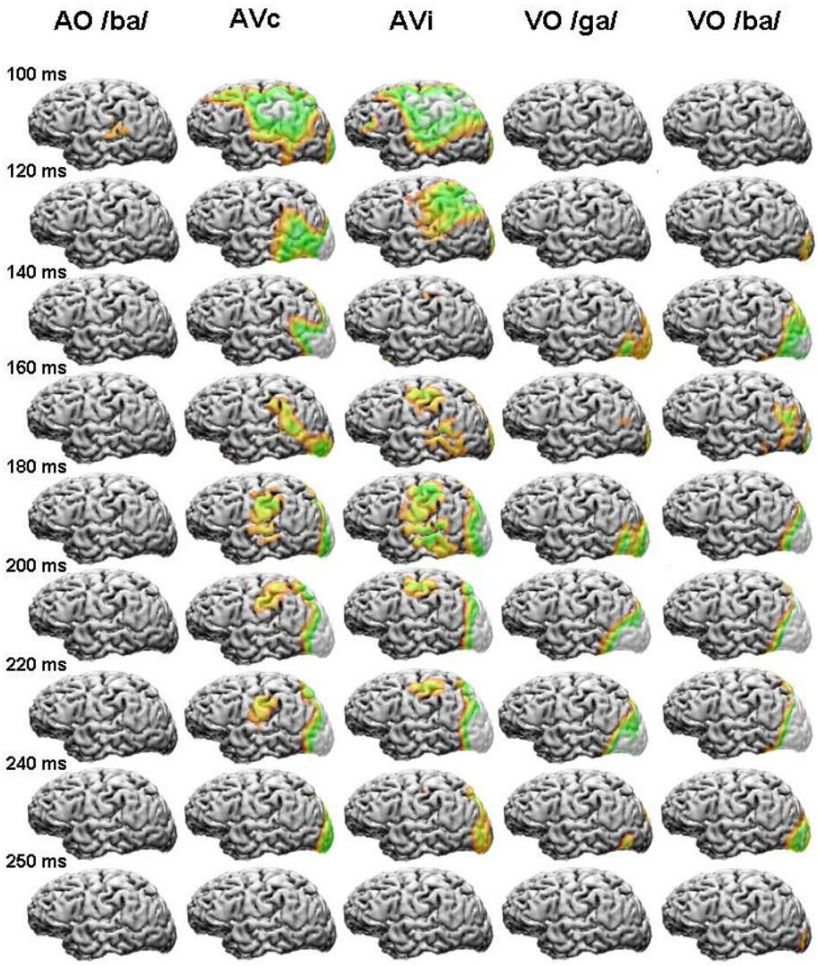


NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

3b

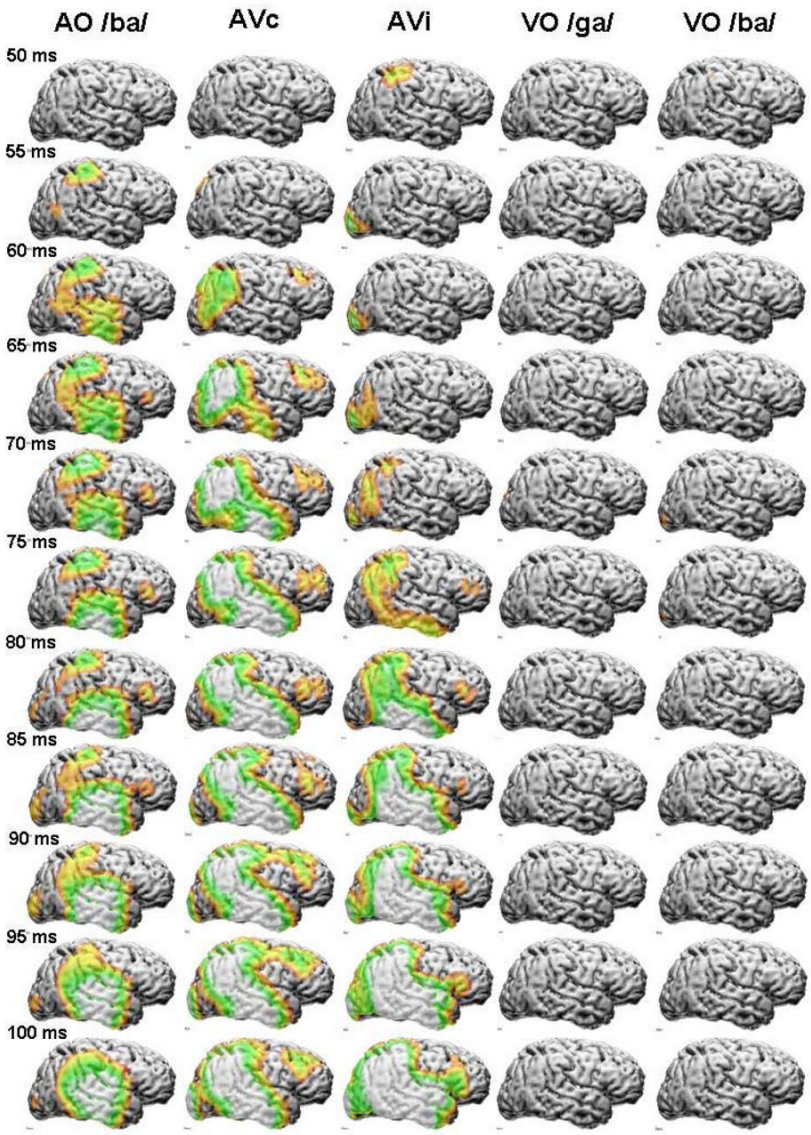


NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

3c



NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

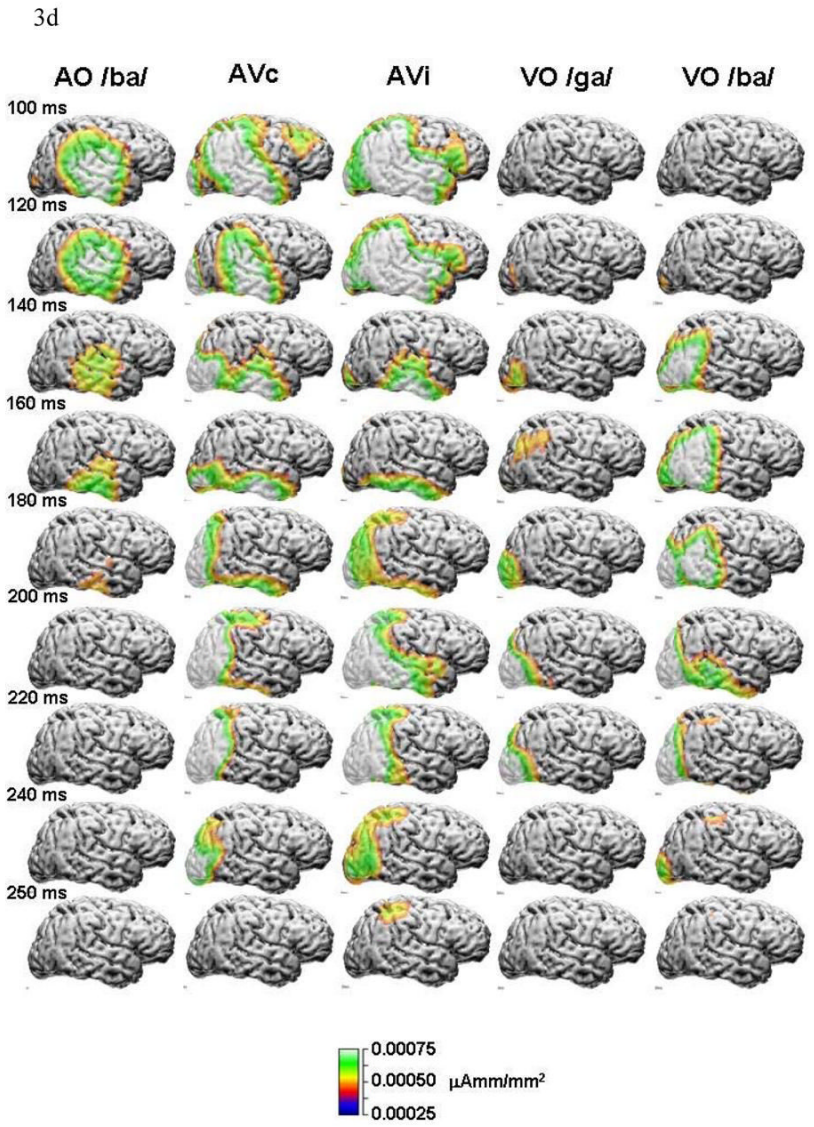


Figure 3(a-d).

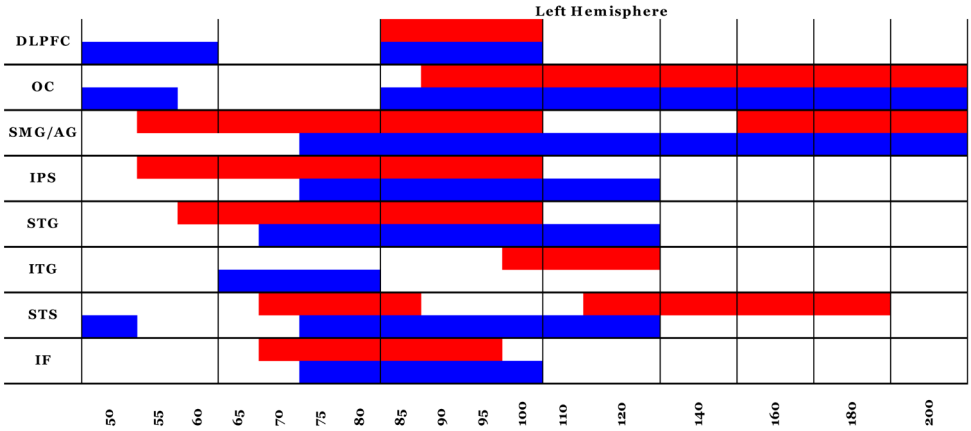
Current density reconstructions for each condition. (a) Left hemisphere from 50 msec to 100 msec in 5-msec steps. Columns are auditory-only (AO) /bα/, audiovisual congruent (AVc) / bα/, audiovisual incongruent (AVi) with auditory /bα/ and visual /gα/, visual-only (VO) /gα/ and VO /bα/. (b) Left hemisphere from 100 to 240 msec in 20-msec steps, with 250 msec, showing that activity was resolved. (c) The right hemisphere from 50 to 100 msec in 5-msec steps. (d) The right hemisphere from 100 to 240 msec in 20-msec steps, with 250 msec, showing that activity was resolved except in the AVi condition.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

4a



4b

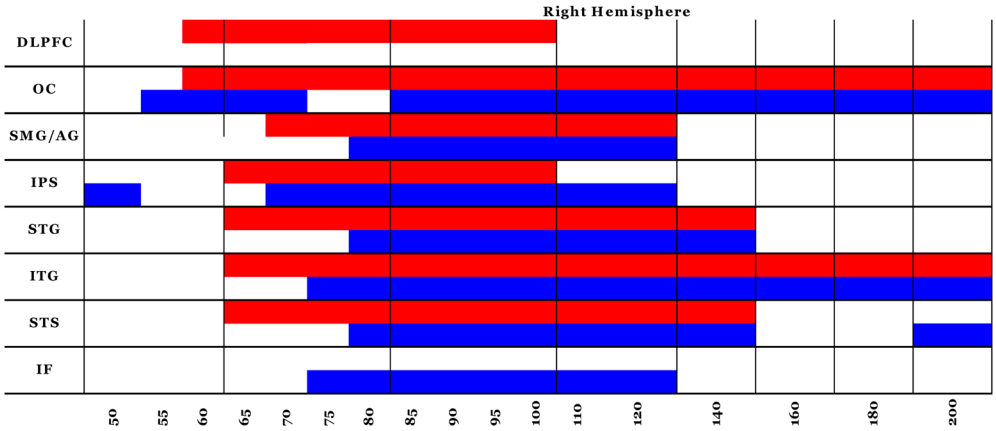


Figure 4(a-b).

Left (4a) and right (4b) hemisphere activation patterns. AVc activity in red, and AVi activity in blue. Time intervals correspond to the resolution in Figure 3(a-d). (Abbreviations: dorsolateral prefrontal cortex, DLPFC; occipital cortex, OC; supramarginal gyrus/angular gyrus, SMG/AG; intraparietal sulcus, IPS; superior temporal gyrus, STG; inferior temporal gyrus, ITG, superior temporal sulcus, STS; inferior frontal, IF).