Cerebral Cortex Advance Access published June 5, 2014

Cerebral Cortex
doi:10.1093/cercor/bhu127

Electrocorticography Reveals Enhanced Visual Cortex Responses to Visual Speech

Inga M. Schepers[1,3], Daniel Yoshor[2] and Michael S. Beauchamp[1]

[1]Department of Neurobiology and Anatomy, University of Texas Medical School at Houston, Houston, TX, USA, [2]Department of Neurosurgery, Baylor College of Medicine, Houston, TX, USA and [3]Current Address: Department of Psychology, Oldenburg University, Oldenburg, Germany

Address correspondence to Email: inga.maren.schepers@uni-oldenburg.de or michael.s.beauchamp@uth.tmc.edu

Human speech contains both auditory and visual components, processed by their respective sensory cortices. We test a simple model in which task-relevant speech information is enhanced during cortical processing. Visual speech is most important when the auditory component is uninformative. Therefore, the model predicts that visual cortex responses should be enhanced to visual-only (V) speech compared with audiovisual (AV) speech. We recorded neuronal activity as patients perceived auditory-only (A), V, and AV speech. Visual cortex showed strong increases in high-gamma band power and strong decreases in alpha-band power to V and AV speech. Consistent with the model prediction, gamma-band increases and alpha-band decreases were stronger for V speech. The model predicts that the uninformative nature of the auditory component (not simply its absence) is the critical factor, a prediction we tested in a second experiment in which visual speech was paired with auditory white noise. As predicted, visual speech with auditory noise showed enhanced visual cortex responses relative to AV speech. An examination of the anatomical locus of the effects showed that all visual areas, including primary visual cortex, showed enhanced responses. Visual cortex responses to speech are enhanced under circumstances when visual information is most important for comprehension.

Keywords: audiovisual, electrocorticography, high gamma, speech, visual cortex

Introduction

Speech is the most important form of human communication and involves both the visual modality (the moving face of the talker) and the auditory modality (the voice of the talker). Simple visual stimuli, such as moving bars, evoke strong responses in visual cortex (Hubel and Wiesel 1968) and these responses can be modulated by simultaneously presented simple auditory stimuli, such as pure tones (Giard and Peronnet 1999; Molholm et al. 2002; Iurilli et al. 2012; Nishimura and Song 2012; Mercier et al. 2013). While visually presented faces also evoke strong visual cortex responses (Pitcher et al. 2011; Davidesco et al. 2013; Schultz et al. 2013) little is known about whether these responses are modulated by auditory stimuli. In particular, visually presented talking faces are usually accompanied by a concomitant auditory stimulus—the voice of the talker.

Speech perception can occur using only information from the auditory modality (e.g., talking on the telephone). When clear auditory speech is present, visual speech information contributes little to speech comprehension (Sumby and Pollack 1954; Bernstein et al. 2004; Ross, Saint-Amour, Leavitt, Javitt et al. 2007). However, when the auditory speech signal is compromised, for example, due to external noise or damage to the cochlea, the visual speech signal gains in importance for comprehension (Bernstein et al. 2004; Ross, Saint-Amour, Leavitt, Javitt et al. 2007; Ross, Saint-Amour, Leavitt, Molholm et al. 2007) and lipreading by itself (speechreading) can be used by deaf individuals to understand the content of speech (Suh et al. 2009).

Because the combination of auditory and visual speech is both ecologically common and important, we hypothesized that visual cortex responses to faces would be strongly modulated by an accompanying voice. Specifically, we test a simple model, which posits that cortical networks for speech perception enhance relevant information and suppress less relevant information. When auditory speech is present (such as during perception of clear AV speech), visual speech information is not necessary for speech perception, and we predict that visual cortex responses should be weak or suppressed. In contrast, under conditions in which no auditory speech information is available (such as visual-only [V] speech) visual information is highly relevant and we predict that the visual cortex responses to visual speech should be strong or enhanced.

Contrary to these predictions, studies using blood oxygen level-dependent functional magnetic resonance imaging (BOLD fMRI) have not reported visual cortex differences in the response to V versus AV speech (Miller and D'Esposito 2005; Wilson et al. 2008; Lee and Noppeney 2011; Nath and Beauchamp 2011; Okada et al. 2013). However, the slow temporal resolution of BOLD fMRI (~2 s for most studies) does not allow for the investigation of the rapid interactions between auditory and visual modalities that occur during speech perception. Therefore, we used electrocorticography (eCog), which allows direct measurement of neural activity with real-time temporal resolution to investigate multisensory interaction in visual cortex during speech perception.

Materials and Methods

Subject Information

Informed consent was obtained from all subjects under the auspices of the Baylor College of Medicine Institutional Review Board and the University of Texas Committees for the Protection of Human Subjects. Seven subjects with medically intractable epilepsy (6 female, mean age 33 years, age range 21–43, 5 right handed according to self-report) participated in the first experiment and 3 different subjects (also with medically intractable epilepsy, one female, ages 23, 36, and 51) participated in the second experiment. Subdural electrodes were implanted in each subject to determine the location of the seizure focus as part of the clinical management of epilepsy, with placement guided solely by clinical criteria. Clinical neurophysiologists identified epileptogenic regions of cortex based on the intracranial recordings. Only data from electrodes that did not exhibit interictal epileptiform activity and that were not found to be sites of seizure onset were analyzed.

Electrode Implantation, Localization, and Recording

Before electrode implantation, structural MR scans were obtained. Cortical surface models were constructed using FreeSurfer (Dale et al. 1999;

© The Author 2014. Published by Oxford University Press. All rights reserved.
For Permissions, please e-mail: journals.permissions@oup.com

Downloaded from http://cercor.oxfordjournals.org/ at Bibliotheks- und Informationssystem der UniversitÃ¤t Oldenburg on June 16, 2014

Fischl et al. 1999) and visualized using SUMA (Argall et al. 2006). After implantation surgery, subjects underwent whole-head CT. The CT scan was aligned to the presurgical structural magnetic resonance (MR) images using the software AFNI and all electrode positions were manually marked on the structural MR images. Subsequently, the electrode positions were assigned to the nearest node on the cortical surface model using the AFNI program SurfaceMetrics. Standard subdural recording electrodes were used (AdTech). eCog was recorded with a 128-channel Blackrock Microsystem (Cerebus, Salt Lake, Utah). The electrodes consisted of platinum alloy discs embedded in silastic with a surface diameter of either 3 mm (regular electrodes) or 0.5 mm (research electrodes); no difference in responses between the electrode types was observed so they were combined for analysis. All visual electrodes were located on strips with inter-electrode distances that varied from 2 to 10 mm.

All electrodes were referenced to an inactive intracranial electrode (an electrode that was turned towards the skull). ECog signals were amplified, online low-pass filtered at 500 Hz (butterworth filter, filter order of 4) and high-pass filtered at 0.3 Hz (butterworth filter, filter order of 1) and digitized at 2 kHz. In the first experiment, we recorded from 650 electrodes in 7 patients. Seventy-one electrodes were classified as visual electrodes, defined anatomically as over occipital lobe and functionally as showing a clear high-gamma band response to visual speech stimulation (averaged response over AV and V conditions). In the second experiment, we recorded from 235 electrodes in 3 patients, selecting 40 visual electrodes.

Experimental Design and Stimuli

During both experiments, subjects were seated in a hospital bed facing a video monitor (Viewsonic VP150, 1024 × 768 pixels) at a viewing distance of 57 cm, with a resulting display size of 40.5 × 22.9°. The presented images covered the entire screen. Sounds were presented from a loudspeaker close to the subject at a volume that was comfortable for the subject and ensured comprehension.

In the first experiment the words DRIVE, KNOWN, LAST, and MEANT (all recorded by the same female talker) were presented in 3 different conditions: V, AV, and auditory-only (A) with 64–96 repetitions per condition. The duration of the stimuli varied between stimuli and conditions (auditory stimuli: DRIVE 500 ms, KNOWN 560 ms, LAST 270 ms, MEANT 420 ms; visual stimuli: DRIVE 1640 ms, KNOWN 1260 ms, LAST 1460 ms, MEANT 1430 ms; AV stimuli had the same duration as the respective visual stimuli). On AV trials, the auditory stimuli started ∼170 ms after the visual stimuli (DRIVE 230 ms, KNOWN 80 ms, LAST 240 ms, MEANT 140 ms). Trials were separated by interstimulus intervals (ISIs) of 2.5 s. On A trials and during the ISIs a fixation dot was shown in the location of the mouth on a gray screen.

In the second experiment 2 additional conditions were introduced: visual speech with auditory white noise ($A_{noise}V$), and A white noise (A_{noise}). In the $A_{noise}V$ condition visual speech input was presented together with auditory white noise that started at the same time as the visual onset and lasted 1.38 s. In the A_{noise} condition white noise was presented for 1.38 s. For each subject, there were 24–80 repetitions (trials) of the 5 stimulus types.

Experimental Task

To ensure attention to the stimuli we used a catch trial design (for one subject in the first experiment, no catch trials were presented). On pseudorandom trials, a catch stimulus was presented, consisting of the AV word PRESS (20% of trials). Subjects were instructed to press a mouse button in response to these trials. The mean accuracy was 89 ± 17% (mean ± SD, for one subject button presses were not recorded) with mean false alarm rate of 3 ± 6%. Instructions were given to look at the fixation dot whenever it was on the screen and at the mouth of the speaker at all other times. Practice trials were used before data collection commenced to ensure that subjects understood the task and were familiar with the stimuli.

Data Analysis

The electrophysiological data were analyzed in MATLAB 7.12.0 (MathWorks Inc. Natick, MA) using the open source toolbox FieldTrip (http://fieldtrip.fcdonferd.nl/, Oostenveld et al. 2011) and customized scripts. The data were low pass filtered at 400 Hz (Butterworth filter, filter order of 4), high pass filtered at 0.5 Hz (Butterworth filter, filter order of 3) and discrete Fourier transform was applied to remove remaining line noise (60, 120, 180 Hz). The data were epoched with respect to the visual stimulus onset from –1 to 2 s.

For the analysis of higher frequencies, the data were transformed to time–frequency space using multitapers (Slepian sequences, number of tapers = 3) followed by Fourier transformation. The tapers were applied in time steps of 10 ms and frequency steps of 2 Hz (frequency smoothing of ±10 Hz, temporal smoothing of 200 ms) to a time window from −0.5 to 1.5 s and a frequency window from 10 to 200 Hz. Before the time–frequency transformations, trials in the A condition were aligned to the visual onset in AV trials. After the time–frequency transformation, the data were inspected at all visual electrodes and single trials were manually removed if they contained clearly visible artifacts. A baseline was calculated over all trials from all experimental conditions at each electrode over the time window from −500 to −200 ms. This baseline was used to calculate the power in percent change at the individual trial level and the data were then averaged over experimental conditions. We first averaged the V and AV responses over all 71 visual electrodes to select a frequency window for the analysis. Based on this response we selected a frequency window from 70 to 110 Hz. To statistically compare the responses to V and AV stimuli at visual electrodes, we averaged the high-gamma band responses over the time window from 0 to 1500 ms after the visual onset and performed a paired *t*-test over all 71 electrodes. To learn more about the temporal dynamics of the difference between the responses, we additionally performed paired *t*-tests between the V and AV condition for each time point (every 10 ms) from 0 to 1500 ms. In the second experiment we performed the same tests between the $A_{noise}V$ and AV condition. The running *t*-tests were false discovery rate (FDR) corrected at $q = 0.05$ (Benjamini and Yekutieli 2001) using the function fdr_bh (Mass Univariate ERP Toolbox; Groppe et al. 2011).

For the analysis of lower frequencies, the data were transformed to time–frequency space using a single Hanning taper with a time window of 500 ms, resulting in a frequency resolution of 2 Hz. The tapers were applied in time steps of 10 ms and frequency steps of 2 Hz to a time window from −0.5 s to 1.5 s and a frequency window from 2 to 30 Hz. A baseline from −500 to −250 ms was used and data analyses were performed in line with the data analyses of high-gamma band activity (see above). A frequency window from 8 to 14 Hz was selected for statistical analyses of lower frequency responses because a clear suppression was seen in this frequency range in the average response over the V and AV conditions.

Anatomy and Response Magnitude

For the first experiment, we classified each visual electrode into one of 4 anatomical regions: visual pole, medial occipital, lateral occipital, and ventral occipital. *T*-tests were performed on the difference responses (V–AV) for each region and an analysis of variance (ANOVA) was performed with region as the fixed factor using the function aov of the R software environment (http://www.r-project.org/). In a separate analysis the geodesic distance of each visual electrode from the occipital pole along the pial surface was obtained using the AFNI SurfDist tool. We calculated the Pearson product-moment correlation and also fit a linear mixed model to the data using the high-gamma difference response as the dependent variable, distance from the occipital pole as the fixed effect and subject as the random effect, using the lmer and lmerTest functions within *R*.

Results

High-Gamma Band Responses to Speech in Visual Cortex

In visual cortex, broadband responses were observed to both AV and V speech (Fig. 1*A*). High-gamma band responses to V speech were much stronger than responses to AV speech (75 ± 53% vs. 63 ± 49%, mean across electrodes ± SD; paired

Downloaded from http://cercor.oxfordjournals.org/ at Bibliotheks- und Informationssystem der UniversitÃ¤t Oldenburg on June 16, 2014

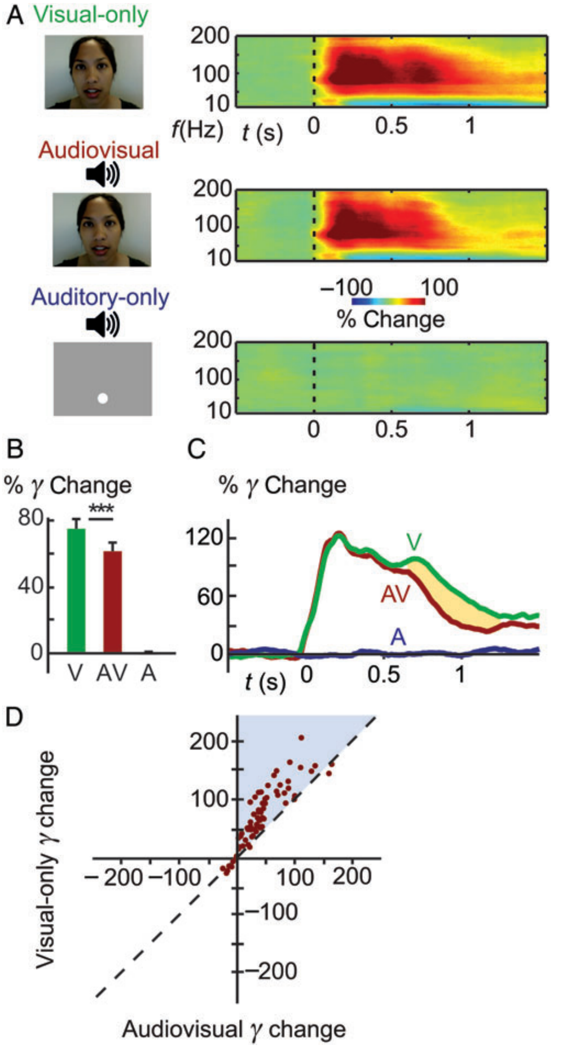

Figure 1. High-frequency responses at visual electrodes. (*A*) Time–frequency responses (10–200 Hz, −0.5–1.5 s) averaged over all 71 visual electrodes for the V (top), AV (middle), and A (bottom) conditions. The dotted lines mark the onset of the visual stimulus (or the respective time point in the A condition). The auditory stimulus onset ∼170 ms later. (*B*) High-gamma-band responses (70–110 Hz, 0–1500 ms) averaged over all visual electrodes for the V, AV, and A conditions (mean ± SEM, *** $P = 10^{-10}$). (*C*) Time courses of the high-gamma band responses (70–110 Hz) across all visual electrodes. The yellow shaded area depicts the significant time window that was used for *D*. (*D*) High-gamma band response to V speech as a function of the AV response magnitude for each visual electrode (70–110 Hz, 660–1230 ms).

t-test: $t_{(70)} = 7.1$, $P = 10^{-10}$; 70–110 Hz, 0–1500 ms; Figure 1*B*). High-gamma band responses to A speech were not significantly above baseline ($0.3 \pm 7\%$, $t_{(70)} = 0.3$, $P = 0.74$).

Next, we investigated the time course of the high-gamma band responses in visual cortex (Fig. 1*C*). The responses rose quickly for both V and AV speech, peaking at 200 ms after the visual onset. There were additional peaks at 400 ms and at 700 ms followed by gradual declines in power between 700 and 1200 ms; the gamma power stayed elevated above baseline until the end of the analysis window at 1500 ms. As expected, the V and AV responses were identical before the onset of the auditory stimulus at 170 ms. The V response was consistently greater than the AV response from 660 to 1230 ms (FDR corrected $P < 0.05$ from 660 to 1230 ms, 1390 to 1400 ms, and 1480 to 1500 ms).

To evaluate the consistency of the greater responses to V speech across visual cortex electrodes, we plotted the magnitude of the V response as a function of the AV response magnitude for each electrode (70–110 Hz, 660–1230 ms; Fig. 1*D*). Most electrodes showed positive gamma-band responses to AV and V speech (AV > 0 and V > 0), and of these, 57/61 (93%) exhibited greater gamma-band increases to V than AV speech. A strong correlation between the V and AV responses across electrodes was observed (Pearson's $r = 0.89$, $P = 10^{-25}$).

Alpha-Band Responses to Speech in Visual Cortex

Our initial analysis focused on high-gamma band activity because it provides an estimate of local neural activity (Ray and Maunsell 2011; Lachaux et al. 2012). We also examined lower frequency responses that have been implicated in speech processing (e.g., Luo and Poeppel 2007; Lange et al. 2013). Unlike the gamma-band *increases* commonly observed in response to sensory stimulation, power in the alpha-band commonly *decreases* in response to sensory stimulation after an initial increase in the evoked response (e.g., Hoogenboom et al. 2006; Siegel et al. 2007; Wyart and Tallon-Baudry 2009; Hipp et al. 2011; Schepers et al. 2012; Davidesco et al. 2013). Consistent with this common finding, we observed strong decreases in the alpha range (~8–14 Hz; Fig. 2*A*) for V and AV speech. However, as in the gamma-band responses, we observed different magnitudes of alpha-band responses for V and AV conditions. The alpha-band decreases were larger for V than AV speech ($-32 \pm 20\%$ vs. $-22 \pm 20\%$, mean across electrodes ± SD; paired *t*-test: $t_{(70)} = -12.4$, $P = 10^{-18}$; 8 to 14 Hz, 0–1500 ms; Figure 2*B*). A speech evoked only weak alpha-band decreases relative to baseline ($-5 \pm 14\%$; $t_{(70)} = -3.2$, $P = 0.002$).

Examining the time course of the alpha-band response revealed an increase during the initial evoked response, peaking at 100 ms after the visual onset. This increase was not different for V and AV speech (Fig. 2*C*). After the initial positive transient, alpha-band responses decreased to below baseline levels, plateauing ∼500 ms after stimulus onset. The alpha-band decrease was larger for the V condition than the AV condition from 550 ms until the end of the analysis window ($p_{\mathrm{FDR}} < 0.05$).

Plotting the alpha-band responses revealed that most individual electrodes showed decreased alpha-band power for V and AV speech (V < 0 and AV < 0; 8–14 Hz, 550–1500 ms; Fig. 2*D*). Of these electrodes, 60/64 (94%) exhibited stronger alpha-band decreases to V than AV speech. There was a strong correlation between V and AV responses ($r = 0.89$, $P = 10^{-24}$).

Experiment 2: Visual Cortex Responses to Speech and Auditory Noise

Compared with AV speech, V speech resulted in greater increases in high-gamma band power and greater decreases in alpha-band power (V > AV), supporting a model in which cortical networks processing task-relevant information are enhanced. Our model predicts that this effect should be specific to meaningful auditory speech information, whose absence triggers an enhanced visual cortex response (resulting in V > AV). However, the effect could also be due to suppression of visual cortex by any simultaneously presented auditory stimulus (also resulting in V > AV). In order to distinguish these possibilities, we conducted a second experiment in which we replaced auditory speech with auditory white noise in the AV condition. If the V > AV response is due to a specific effect of

Downloaded from http://cercor.oxfordjournals.org/ at Bibliotheks- und Informationssystem der UniversitÃ¤t Oldenburg on June 16, 2014

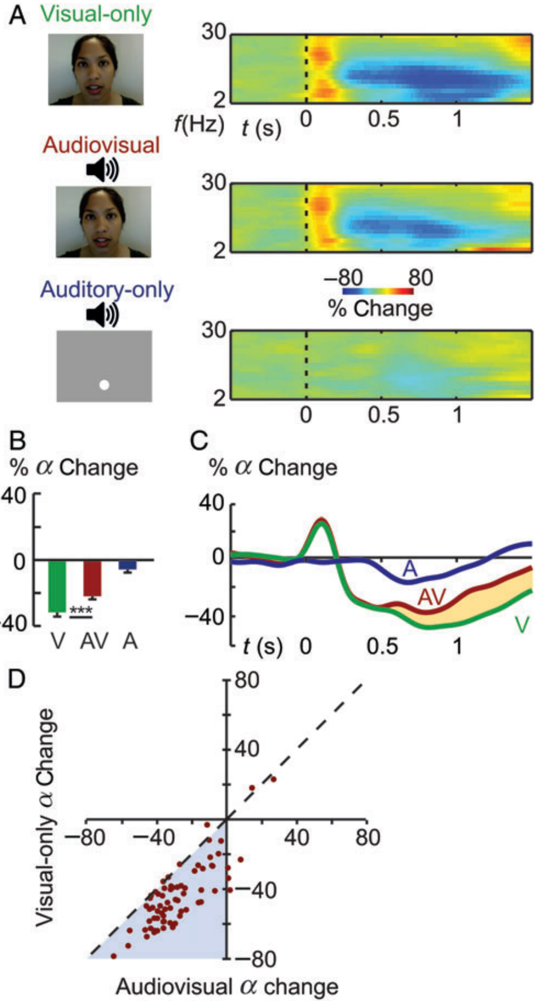

Figure 2. Low-frequency responses at visual electrodes. (*A*) Time–frequency responses (2–30 Hz, −0.5 to 1.5 s) averaged over all 71 visual electrodes for the V (top), AV (middle), and A (bottom) conditions. The dotted lines mark the onset of the visual stimulus (or the respective time point in the A condition). The auditory stimulus onset ∼170 ms later. (*B*) Alpha-band responses (8–14 Hz, 0–1500 ms) averaged over all visual electrodes for the V, AV, and A conditions (mean ± SEM, *** $P = 10^{-18}$). (*C*) Time courses of the alpha-band responses (8–14 Hz) across all visual electrodes. The yellow shaded area depicts the significant time window that was used for *D*. (*D*) Alpha-band response to V speech as a function of the AV response magnitude for each visual electrode (8–14 Hz, 550–1500 ms).

meaningful auditory speech, then auditory noise should be similar to no sound at all: V ≅ A_{noise}V and A_{noise}V > AV. In contrast, if the V > AV response is due to a nonspecific suppression of visual cortex by sound, then auditory noise should produce a similar effect: V > A_{noise}V and A_{noise}V ≅ AV.

To distinguish these possibilities, V, AV, and A speech were presented, along with 2 additional conditions in which the auditory speech in the original stimuli was replaced with auditory white noise: visual speech with auditory white noise (A_{noise}V) and A white noise (A_{noise}). Because the visual components of AV and A_{noise}V were identical, any differences in visual cortex responses must be due to the difference between auditory meaningful speech and white noise.

First, we replicated the result from the first experiment. High-gamma band responses to V speech were significantly greater than responses to AV speech (142 ± 90% vs. 108 ± 66%, mean across electrodes ± SD; paired *t*-test: $t_{(29)} = 3.8$, $P = 10^{-4}$; 70–110 Hz, 0–1500 ms) and alpha-band decreases were greater to V than AV speech (−38 ± 16% vs. −27 ± 17%, mean across electrodes ± SD; paired *t*-test: $t_{(29)} = -4.4$, $P = 10^{-5}$; 8–14 Hz, 0–1500 ms).

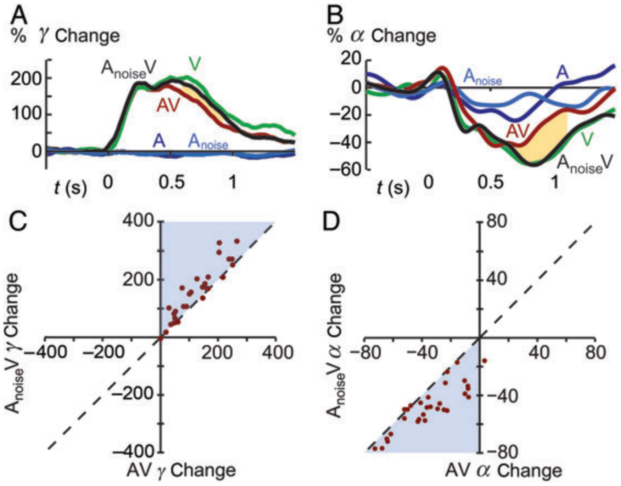

Figure 3. Results of the second experiment. (*A*) Time courses of the high-gamma band responses (70–110 Hz) across all electrodes for the AV noise (black), AV speech (red), V (green), A (dark blue), and auditory noise (light blue) conditions. The yellow shaded area depicts the significant time window that was used for C. (*B*) Time courses of the alpha-band responses (8–14 Hz). (*C*) High-gamma band response to AV noise as a function of the AV response magnitude for each visual electrode (70–110 Hz, 610–920 ms). (*D*) Alpha-band response to AV noise as a function of the AV response magnitude for each visual electrode (8–14 Hz, 720–1100 ms).

Next, we compared gamma-band responses to A_{noise}V and AV speech. Responses to A_{noise}V speech were stronger than the responses to AV speech (Fig. 3*A*). The responses began to diverge at 610 ms, with the A_{noise}V response greater than the AV response from 610 to 920 ms (all pFDR < 0.05). Plotting the individual electrode responses (70–110 Hz, 610–920 ms; Fig. 3*C*) showed that most electrodes displayed positive gamma-band responses to both stimulus types (A_{noise}V > 0 and AV > 0). Of these electrodes, 23/28 (82%) exhibited greater increases to A_{noise}V than AV.

Finally, we compared the alpha-band responses. There were greater alpha-band decreases for A_{noise}V compared with AV speech (Fig. 3*B*). The effect was significant from 140 to 290 ms, 720 to 1100 ms and 1450 to 1500 ms (all pFDR < 0.05). Most individual electrodes showed alpha-band decreases (8–14 Hz, 720–1100 ms; Fig. 3*D*) for both stimulus types (A_{noise}V < 0 and AV < 0); of these, 26/29 (90%) exhibited greater alpha-band decreases to A_{noise}V than AV.

In summary, the visual cortex responses to visual speech with auditory noise and no auditory component were similar, suggesting that the lack of a meaningful auditory stimulus is critical.

Anatomical Specificity of Visual Cortex Responses

In both experiments, we observed significantly greater visual cortex responses to V speech than AV speech. If the visual cortex enhancement was mediated by direct projections from auditory cortex, early visual areas might be expected to show stronger enhancement because of the documented projections from auditory cortex to striate cortex (Falchier et al. 2002; Nishimura and Song 2012). Alternatively, if visual cortex suppression was mediated by parietal or frontal areas, later visual

Downloaded from http://cercor.oxfordjournals.org/ at Bibliotheks- und Informationssystem der UniversitÃ¤t Oldenburg on June 16, 2014

areas might show a larger effect because of their stronger connections with higher cortical areas (Lewis and van Essen 2000). We created cortical surface models of each individual subject and mapped the activity from each individual electrode on to the cortical surface model, with the magnitude of the difference between the V and AV responses assigned to a color scale (Fig. 4).

Significantly greater gamma-band responses to V than AV speech were observed at all locations in visual cortex, including cortex on the banks of the calcarine sulcus, the location of primary visual cortex (V1). However, there was no obvious gradient in the response difference between different visual areas. To quantify this observation, we divided the electrodes by their anatomical location into occipital pole (OP, $n = 15$), medial occipital (MO, $n = 34$), lateral occipital (LO, $n = 18$), and ventral occipital (VO, $n = 4$). While all regions showed enhanced high-gamma band responses to V speech (70–110 Hz, 0–1500 ms; OP: $t_{(14)} = 3.6$, $P = 0.003$; MO: $t_{(33)} = 3.5$, $P = 10^{-4}$; LO: $t_{(17)} = 6.2$, $P = 10^{-6}$; VO: $t_{(3)} = 2.3$, $P = 0.105$; Fig. 4B), an analysis of variance showed no differences between regions ($F_{3,210} < 0.1$, $P > 0.9$). We also measured the distance of each electrode in each subject from the occipital pole (Murphey et al. 2009) and found a similar lack of anatomical specificity. There was no dependence between distance and the difference responses with either a linear correlation ($r = -0.04$, $P = 0.8$) or a linear mixed effect model with distance as fixed effect and subject as random effect (estimate $= -0.04$, $P = 0.68$).

A similar lack of effect of anatomical location was observed in the alpha-band. On the cortical surface model, stronger alpha-band decreases for V than AV speech were observed at all locations in visual cortex (Fig. 4C). Classifying electrodes by region showed consistently greater alpha-band decreases to V

Downloaded from http://cercor.oxfordjournals.org/ at Bibliotheks- und Informationssystem der UniversitÃ¤t Oldenburg on June 16, 2014

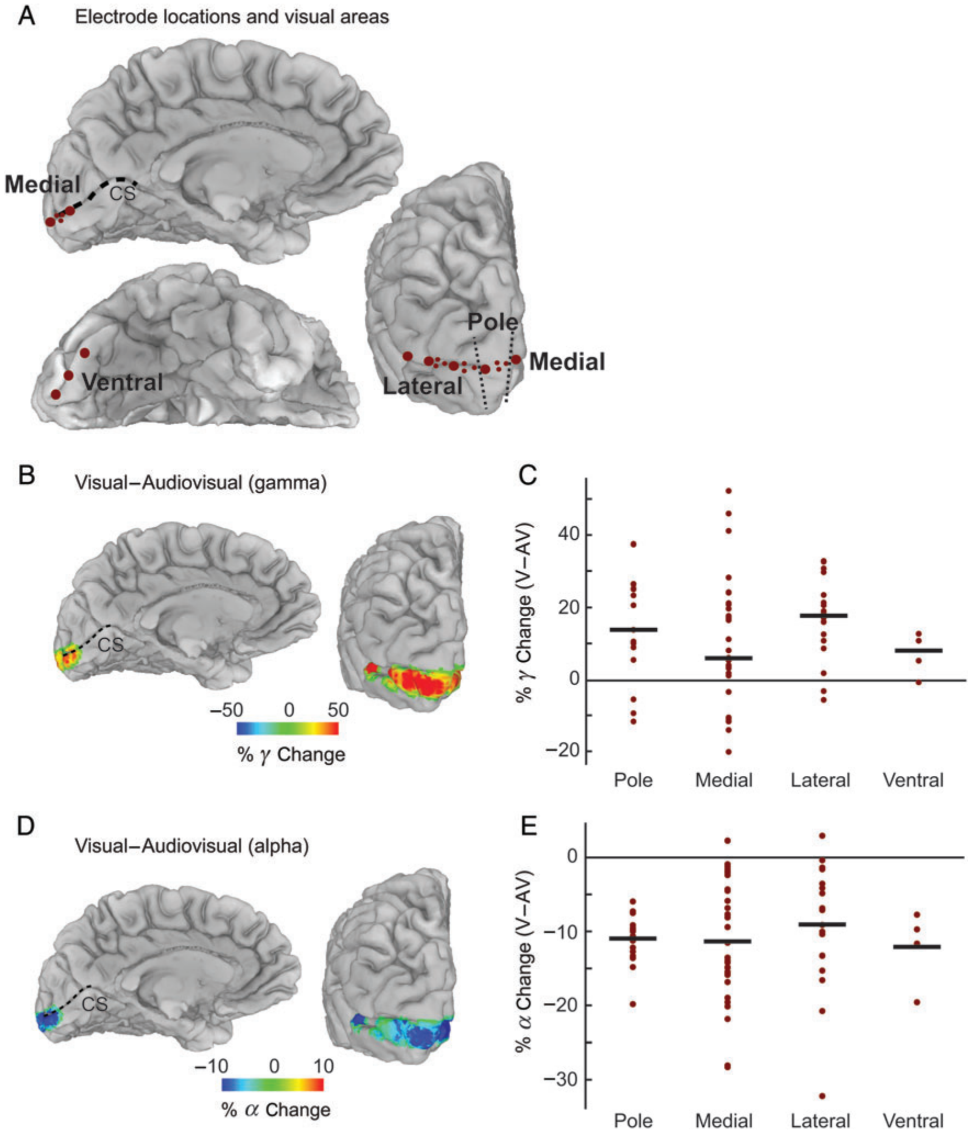

Figure 4. Electrode coverage and classification. (*A*) Electrode locations over visual cortex in 2 representative subjects: YAH (medial and posterior views) and YAD (ventral view). Electrodes are shown as burgundy circles. Each electrode was classified by anatomical region, dotted black line shows boundary between regions (Pole: occipital pole; CS and dashed black line: calcarine sulcus). (*B*) Difference in high-gamma responses (V–AV) for each electrode in subject YAH, mapped to the cortical surface. (*C*) Difference in high-gamma responses (V–AV) for all electrodes, sorted by anatomical region. Black bar shows mean for each region; no effect of region was observed. (*D*) Difference in alpha responses (V–AV) in subject YAH. (*E*) Difference in alpha responses (V–AV) for all electrodes, sorted by anatomical region. No effect of region was observed.

speech (8–14 Hz, 0–1500 ms; OP: $t_{(14)} = -12.5$, $P = 10^{-8}$; MO: $t_{(33)} = -7.9$, $P = 10^{-8}$; LO: $t_{(17)} = -4.7$, $P = 10^{-3}$; VO: $t_{(3)} = -4.7$, $P = 0.018$; Fig. 4*D*), but an analysis of variance showed no differences between regions ($F_{3,210} = 0.3$, $P = 0.8$). There was no dependence between distance and the difference responses with either a linear correlation ($r = -0.06$, $P = 0.6$) or a linear mixed effect model with distance as fixed effect and subject as random effect (estimate = −0.05, $P = 0.4$).

Discussion

Using eCog, we examined responses to multisensory speech with high spatial and temporal resolution. In both early and late visual cortex, we observed greater responses to visual speech than AV speech, suggesting an enhancement when visual speech is presented by itself. A second experiment showed that this enhancement remained when auditory noise was presented, indicating that the meaningful content of the auditory stimulus was necessary to induce the observed response difference between V and AV speech.

Previous studies have observed both low-frequency decreases and high-gamma band increases to face stimuli in visual cortex, as observed in the present study (Kaiser et al. 2005; Dobel et al. 2011; Davidesco et al. 2013; Grützner et al. 2013; Schepers et al. 2013). However, previous studies have not compared the difference between the responses to V and AV speech. For instance, in an electroencephalography (EEG) study, AV speech and A speech both evoked occipital decreases in low-frequency power (Schepers et al. 2013), but because V speech was not presented, enhanced responses to V versus AV speech could not be observed.

In addition to the sustained low-frequency decreases and high-gamma band increases that were the most prominent feature of our data, we also observed a low-frequency response increase that peaked at ~100 ms, similar to the event-related potential (ERP) responses reported in both scalp EEG and magnetoencephalography (MEG) studies of face processing (Halgren et al. 2000; Liu et al. 2002; Feng et al. 2012; Wu et al. 2012) and eCog recordings (Allison et al. 1999). We saw no difference in these early responses between visual and AV speech; this is unsurprising because the onset of auditory speech did not occur until ~170 ms after visual onset. This temporal delay between the onset of the multisensory interaction and the timing of the short latency transient response may explain why previous studies did not report differences in visual cortex ERP responses to V versus AV speech (Besle et al. 2004).

fMRI Studies on AV Speech

Neuroimaging studies have reported strong responses in early and late visual cortex to AV and visual speech (Miller and D'Esposito 2005; Wilson et al. 2008; Lee and Noppeney 2011; Nath and Beauchamp 2011; Okada et al. 2013) without finding differences in the response to V versus AV speech. This disparity between the fMRI and eCog results may be due to the relatively slow temporal resolution of fMRI. In our results, the difference between responses to V versus AV speech was most pronounced at 890 ms after visual onset. Because fMRI integrates responses over the relatively long duration of each repetition time (TR–2 s for most studies) it may lack sensitivity to responses or response differences that occur in restricted time windows. For instance, visually presented faces that are not attended by the subject evoke a short-duration response that is detected with MEG but not with fMRI (Furey et al. 2006). With high temporal resolution fMRI (effective TR–100 ms), it may be possible to replicate the findings described in the present study (Lin et al. 2012; Chang et al. 2013).

Source of the Observed Effects in Visual Cortex

As shown in Figure 5, perception of AV speech activates an extended network of brain areas consisting of visual cortex, auditory cortex, and multisensory areas in the superior temporal sulcus (STS) as well as parietal and frontal areas (Price et al. 2005; Hickok and Poeppel 2007; Pulvermuller and Fadiga 2010; Nath and Beauchamp 2011). We suggest a simple model in which the goal of this network is to extract meaning from AV speech as quickly and efficiently as possible. When the auditory component of the speech input is sufficient to extract meaning, the visual response is not needed and therefore not enhanced. In contrast, when the auditory stimulus does not contain sufficient information to extract meaning (as is the case for V speech or visual speech with auditory white noise) the visual response is needed to extract meaning and is therefore enhanced.

Because of its long latency and context dependence, the enhancement of visual-cortex responses likely relies on top-down neural circuitry consisting of recurrent connections from frontal-parietal-temporal areas to visual areas, perhaps including thalamic relays (Sherman 2007; Saalmann et al. 2012; Davidesco et al. 2013; Leitao et al. 2013). The speech network may engage the top-down circuitry without requiring a conscious (voluntary) shift of attention. In this model, auditory association areas perform initial processing of the auditory speech stimulus, as early as 110–150 ms after auditory stimulus onset in the superior temporal gyrus (Chang et al. 2010; MacGregor et al. 2012). If the auditory speech signal is not sufficient to unambiguously identify the speech stimulus, higher-level areas would become engaged to further process the auditory and visual speech information, including dorsolateral prefrontal cortex, the frontal eye field and lateral intraparietal cortex (Rossi et al. 2007; Bisley and Goldberg 2010; Miller and Buschman 2013; Squire et al. 2013) and STS (Beauchamp et al. 2004; Kayser and Logothetis 2009). Regardless of the anatomical origin of the visual enhancement, the model predicts that presenting even an irrelevant visual stimulus at the location of

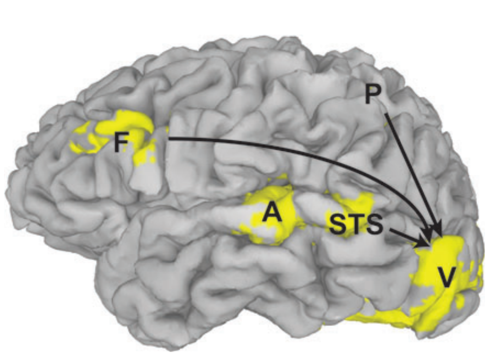

Figure 5. Regions important for cortical AV speech processing. Cortical surface model showing BOLD fMRI responses (yellow color) to AV speech ($P < 0.05$, FDR corrected; subject KO). F, frontal areas; P, parietal areas (activity in depth of intraparietal sulcus not visible), A, auditory cortex; STS, superior temporal sulcus and related multisensory areas; V, visual cortex. Arrows indicate possible pathways for enhancement of visual cortex activity when auditory information is insufficient for speech comprehension.

Downloaded from http://cercor.oxfordjournals.org/ at Bibliotheks- und Informationssystem der UniversitÃ¤t Oldenburg on June 16, 2014

the talker's face would evoke a larger response in the V condition than the AV condition. The model also predicts that presenting AV speech with increasing degrees of auditory noise would lead to graded enhancement of visual responses (Kawase et al. 2005; Nath and Beauchamp 2011; Schepers et al. 2013).

We observed a consistent difference in visual cortex responses to V versus AV speech. We refer to this difference as an enhanced response to V speech relative to AV speech because in normal experience, AV speech is encountered far more often than V speech and it therefore seems sensible to consider it as the baseline condition. Of course, the effect could also be considered a diminished response to AV speech relative to V speech. This would suggest that auditory cortex inhibits visual cortex, leading to a diminished response in visual cortex when the auditory component of the stimulus is present in AV speech. This direct inhibition would be expected to occur early after auditory stimulation. For instance, an auditory tone suppressed visual responses to a disc beginning at ~40 ms after stimulus onset (Giard and Peronnet 1999; Molholm et al. 2002; Mercier et al. 2013). In mice, activation of auditory cortex through short noise bursts let to reduced visual neural responses to moving bars in primary visual cortex beginning ~6 ms after auditory stimulus onset (Iurilli et al. 2012). Anatomically, these effects could be mediated by direct axonal projections from auditory cortex to visual cortex or thalamocortical connections (Falchier et al. 2002; Campi et al. 2010; van den Brink et al. 2013).

There are 2 arguments against such direct inhibition of visual cortex by auditory cortex as the explanation for the effects observed in our study. The first is the long latency of the observed response difference between V versus AV speech: it did not begin until ~490 ms after auditory stimulus onset, much later than that reported in studies that used simple stimuli. The second is that the inhibition was contingent upon the content of the auditory stimulus: in the second experiment, auditory noise paired with visual speech showed greater responses than AV speech.

Implications for Sensory Deficits

Hearing loss occurs in ~80% of people with healthy aging and auditory word recognition declines in healthy older adults (Gates et al. 1990; Yueh et al. 2003). Interestingly, older adults with hearing impairments have been found to perform better in word identification than older adults with normal hearing when only the visual speech signal was available (Tye-Murray et al. 2007). When the auditory speech signal is compromised, for example, due to external noise or damage to the cochlea, the visual speech signal gains in importance for comprehension (Bergeson et al. 2005; Ross, Saint-Amour, Leavitt, Javitt et al. 2007). Therefore, patients with impaired auditory perception might be expected to have heightened visual cortex responses to speech.

Funding

This work was supported by the National Institutes of Health (R01NS065395) and the German Research Foundation (SCHE 1815/1-1).

Notes

We thank the subjects and their families and the clinical staff at St. Luke''s Episcopal Hospital, whose cooperation made this research possible. We thank Christen Symank and Debshila Basu Mallick for assistance with MR data collection and Xiaomei Pei and Ping Sun for assistance with eCog data collection.

References

Allison T, Puce A, Spencer DD, McCarthy G. 1999. Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. Cereb Cortex. 9:415–430.

Argall BD, Saad ZS, Beauchamp MS. 2006. Simplified intersubject averaging on the cortical surface using SUMA. Hum Brain Mapp. 27:14–27.

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.

Benjamini Y, Yekutieli D. 2001. The control of the false discovery rate in multiple testing under dependency. Ann Stat. 29:1165–1188.

Bergeson TR, Pisoni DB, Davis RA. 2005. Development of audiovisual comprehension skills in prelingually deaf children with cochlear implants. Ear Hear. 26:149–164.

Bernstein LE, Auer JET, Takayanagi S. 2004. Auditory speech detection in noise enhanced by lipreading. Speech Commun. 44:5–18.

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

Bisley JW, Goldberg ME. 2010. Attention, intention, and priority in the parietal lobe. Annu Rev Neurosci. 33:1–21.

Campi KL, Bales KL, Grunewald R, Krubitzer L. 2010. Connections of auditory and visual cortex in the prairie vole (*Microtus ochrogaster*): evidence for multisensory processing in primary sensory areas. Cereb Cortex. 20:89–108.

Chang EF, Rieger JW, Johnson K, Berger MS, Barbaro NM, Knight RT. 2010. Categorical speech representation in human superior temporal gyrus. Nat Neurosci. 13:1428–1432.

Chang WT, Nummenmaa A, Witzel T, Ahveninen J, Huang S, Tsai KW, Chu YH, Polimeni JR, Belliveau JW, Lin FH. 2013. Whole-head rapid fMRI acquisition using echo-shifted magnetic resonance inverse imaging. Neuroimage. 78:325–338.

Dale AM, Fischl B, Sereno MI. 1999. Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage. 9:179–194.

Davidesco I, Harel M, Ramot M, Kramer U, Kipervasser S, Andelman F, Neufeld MY, Goelman G, Fried I, Malach R. 2013. Spatial and object-based attention modulates broadband high-frequency responses across the human visual cortical hierarchy. J Neurosci. 33:1228–1240.

Dobel C, Junghofer M, Gruber T. 2011. The role of gamma-band activity in the representation of faces: reduced activity in the fusiform face area in congenital prosopagnosia. PLoS ONE. 6:e19550.

Falchier A, Clavagnier S, Barone P, Kennedy H. 2002. Anatomical evidence of multimodal integration in primate striate cortex. J Neurosci. 22:5749–5759.

Feng W, Martinez A, Pitts M, Luo YJ, Hillyard SA. 2012. Spatial attention modulates early face processing. Neuropsychologia. 50: 3461–3468.

Fischl B, Sereno MI, Dale AM. 1999. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. Neuroimage. 9:195–207.

Furey ML, Tanskanen T, Beauchamp MS, Avikainen S, Uutela K, Hari R, Haxby JV. 2006. Dissociation of face-selective cortical responses by attention. Proc Natl Acad Sci USA. 103:1065–1070.

Gates GA, Cooper JC Jr, Kannel WB, Miller NJ. 1990. Hearing in the elderly: the Framingham cohort, 1983–1985. Part I. Basic audiometric test results. Ear Hear. 11:247–256.

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.

Groppe DM, Urbach TP, Kutas M. 2011. Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. Psychophysiology. 48:1711–1725.

Downloaded from http://cercor.oxfordjournals.org/ at Bibliotheks- und Informationssystem der UniversitÃ¤t Oldenburg on June 16, 2014

Grützner C, Wibral M, Sun L, Rivolta D, Singer W, Maurer K, Uhlhaas PJ. 2013. Deficits in high- (>60 Hz) gamma-band oscillations during visual processing in schizophrenia. Front Hum Neurosci. 7:88.

Halgren E, Raij T, Marinkovic K, Jousmaki V, Hari R. 2000. Cognitive response profile of the human fusiform face area as determined by MEG. Cereb Cortex. 10:69–81.

Hickok G, Poeppel D. 2007. The cortical organization of speech processing. Nat Rev Neurosci. 8:393–402.

Hipp JF, Engel AK, Siegel M. 2011. Oscillatory synchronization in large-scale cortical networks predicts perception. Neuron. 69:387–396.

Hoogenboom N, Schoffelen JM, Oostenveld R, Parkes LM, Fries P. 2006. Localizing human visual gamma-band activity in frequency, time and space. Neuroimage. 29:764–773.

Hubel DH, Wiesel TN. 1968. Receptive fields and functional architecture of monkey striate cortex. J Physiol. 195:215–243.

Iurilli G, Ghezzi D, Olcese U, Lassi G, Nazzaro C, Tonini R, Tucci V, Benfenati F, Medini P. 2012. Sound-driven synaptic inhibition in primary visual cortex. Neuron. 73:814–828.

Kaiser J, Hertrich I, Ackermann H, Mathiak K, Lutzenberger W. 2005. Hearing lips: gamma-band activity during audiovisual speech perception. Cereb Cortex. 15:646–653.

Kawase T, Yamaguchi K, Ogawa T, Suzuki K, Suzuki M, Itoh M, Kobayashi T, Fujii T. 2005. Recruitment of fusiform face area associated with listening to degraded speech sounds in auditory-visual speech perception: a PET study. Neurosci Lett. 382:254–258.

Kayser C, Logothetis NK. 2009. Directed interactions between auditory and superior temporal cortices and their role in sensory integration. Front Integr Neurosci. 3:7.

Lachaux JP, Axmacher N, Mormann F, Halgren E, Crone NE. 2012. High-frequency neural activity and human cognition: past, present and possible future of intracranial EEG research. Prog Neurobiol. 98:279–301.

Lange J, Christian N, Schnitzler A. 2013. Audio-visual congruency alters power and coherence of oscillatory activity within and between cortical areas. Neuroimage. C79:111–120.

Lee H, Noppeney U. 2011. Physical and perceptual factors shape the neural mechanisms that integrate audiovisual signals in speech comprehension. J Neurosci. 31:11338–11350.

Leitao J, Thielscher A, Werner S, Pohmann R, Noppeney U. 2013. Effects of parietal TMS on visual and auditory processing at the primary cortical level—a concurrent TMS-fMRI study. Cereb Cortex. 23:873–884.

Lewis JW, van Essen DC. 2000. Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. J Comp Neurol. 428:112–137.

Lin FH, Witzel T, Raij T, Ahveninen J, Tsai KW, Chu YH, Chang WT, Nummenmaa A, Polimeni JR, Kuo WJ et al. 2012. fMRI hemodynamics accurately reflects neuronal timing in the human brain measured by MEG. Neuroimage. 78:372–384.

Liu J, Harris A, Kanwisher N. 2002. Stages of processing in face perception: an MEG study. Nat Neurosci. 5:910–916.

Luo H, Poeppel D. 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron. 54:1001–1010.

MacGregor LJ, Pulvermuller F, van Casteren M, Shtyrov Y. 2012. Ultra-rapid access to words in the brain. Nat Commun. 3:711.

Mercier MR, Foxe JJ, Fiebelkorn IC, Butler JS, Schwartz TH, Molholm S. 2013. Auditory-driven phase reset in visual cortex: human electrocorticography reveals mechanisms of early multisensory integration. Neuroimage. 79C:19–29.

Miller EK, Buschman TJ. 2013. Cortical circuits for the control of attention. Curr Opin Neurobiol. 23:216–222.

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

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

Murphey DK, Maunsell JH, Beauchamp MS, Yoshor D. 2009. Perceiving electrical stimulation of identified human visual areas. Proc Natl Acad Sci USA. 106:5389–5393.

Nath AR, Beauchamp MS. 2011. Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. J Neurosci. 31:1704–1714.

Nishimura M, Song WJ. 2012. Temporal sequence of visuo-auditory interaction in multiple areas of the guinea pig visual cortex. PLoS ONE. 7:e46339.

Okada K, Venezia JH, Matchin W, Saberi K, Hickok G. 2013. An fMRI study of audiovisual speech perception reveals multisensory interactions in auditory cortex. PLoS ONE. 8(6):68–959.

Oostenveld R, Fries P, Maris E, Schoffelen JM. 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci. 2011:156869.

Pitcher D, Dilks DD, Saxe RR, Triantafyllou C, Kanwisher N. 2011. Differential selectivity for dynamic versus static information in face-selective cortical regions. Neuroimage. 56:2356–2363.

Price C, Thierry G, Griffiths T. 2005. Speech-specific auditory processing: where is it? Trends Cogn Sci. 9:271–276.

Pulvermuller F, Fadiga L. 2010. Active perception: sensorimotor circuits as a cortical basis for language. Nat Rev Neurosci. 11:351–360.

Ray S, Maunsell JH. 2011. Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. PLoS Biol. 9:e1000610.

Ross LA, Saint-Amour D, Leavitt VM, Javitt DC, Foxe JJ. 2007. Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. Cereb Cortex. 17:1147–1153.

Ross LA, Saint-Amour D, Leavitt VM, Molholm S, Javitt DC, Foxe JJ. 2007. Impaired multisensory processing in schizophrenia: deficits in the visual enhancement of speech comprehension under noisy environmental conditions. Schizophr Res. 97:173–183.

Rossi AF, Bichot NP, Desimone R, Ungerleider LG. 2007. Top down attentional deficits in macaques with lesions of lateral prefrontal cortex. J Neurosci. 27:11306–11314.

Saalmann YB, Pinsk MA, Wang L, Li X, Kastner S. 2012. The pulvinar regulates information transmission between cortical areas based on attention demands. Science. 337:753–756.

Schepers IM, Hipp JF, Schneider TR, Roder B, Engel AK. 2012. Functionally specific oscillatory activity correlates between visual and auditory cortex in the blind. Brain. 135:922–934.

Schepers IM, Schneider TR, Hipp JF, Engel AK, Senkowski D. 2013. Noise alters beta-band activity in superior temporal cortex during audiovisual speech processing. Neuroimage. 70:101–112.

Schultz J, Brockhaus M, Bulthoff HH, Pilz KS. 2013. What the human brain likes about facial motion. Cereb Cortex. 23:1167–1178.

Sherman SM. 2007. The thalamus is more than just a relay. Curr Opin Neurobiol. 17:417–422.

Siegel M, Donner TH, Oostenveld R, Fries P, Engel AK. 2007. High-frequency activity in human visual cortex is modulated by visual motion strength. Cereb Cortex. 17:732–741.

Squire RF, Noudoost B, Schafer RJ, Moore T. 2013. Prefrontal contributions to visual selective attention. Annu Rev Neurosci. 36:451–466.

Suh MW, Lee HJ, Kim JS, Chung CK, Oh SH. 2009. Speech experience shapes the speechreading network and subsequent deafness facilitates it. Brain. 132:2761–2771.

Sumby WH, Pollack I. 1954. Visual contribution to speech intelligibility in noise. J Acoust Soc Am. 26:212–215.

Tye-Murray N, Sommers MS, Spehar B. 2007. Audiovisual integration and lipreading abilities of older adults with normal and impaired hearing. Ear Hear. 28:656–668.

van den Brink RL, Cohen MX, van der Burg E, Talsma D, Vissers ME, Slagter HA. 2013. Subcortical, modality-specific pathways contribute to multisensory processing in humans. Cereb Cortex. doi: 10.1093/cercor/bht069.

Wilson SM, Molnar-Szakacs I, Iacoboni M. 2008. Beyond superior temporal cortex: intersubject correlations in narrative speech comprehension. Cereb Cortex. 18:230–242.

Wu J, Duan H, Tian X, Wang P, Zhang K. 2012. The effects of visual imagery on face identification: an ERP study. Front Hum Neurosci. 6:305.

Wyart V, Tallon-Baudry C. 2009. How ongoing fluctuations in human visual cortex predict perceptual awareness: baseline shift versus decision bias. J Neurosci. 29:8715–8725.

Yueh B, Shapiro N, MacLean CH, Shekelle PG. 2003. Screening and management of adult hearing loss in primary care: scientific review. JAMA. 289:1976–1985.

Downloaded from http://cercor.oxfordjournals.org/ at Bibliotheks- und Informationssystem der UniversitÃ¤t Oldenburg on June 16, 2014