Published in final edited form as:

Neurosci Lett. 2009 March 20; 452(3): 219–223. doi:10.1016/j.neulet.2009.01.060.

Two cortical mechanisms support the integration of visual and auditory speech: A hypothesis and preliminary data

Kayoko Okada and Gregory Hickok

Department of Cognitive Sciences, University of California, Irvine, Irvine, CA 92697-3800

Abstract

Visual speech (lip-reading) influences the perception of heard speech. The literature suggests at least two possible mechanisms for this influence: “direct” sensory-sensory interaction, whereby sensory signals from auditory and visual modalities are integrated directly, likely in the superior temporal sulcus, and “indirect” sensory-motor interaction, whereby visual speech is first mapped onto motor-speech representations in the frontal lobe, which in turn influences sensory perception via sensory-motor integration networks. We hypothesize that both mechanisms exist, and further that lip-reading functional activations of Broca’s region and the posterior planum temporale reflect the sensory-motor mechanism. We tested one prediction of this hypothesis using fMRI. We assessed whether viewing visual speech (contrasted with facial gestures) activates the same network as a speech sensory-motor integration task (listen to and then silently rehearse speech). Both tasks activated locations within Broca’s area, dorsal premotor cortex, and the posterior planum temporal (Spt), and focal regions of the STS, all of which have previously been implicated in sensory-motor integration for speech. This finding is consistent with the view that visual speech influences heard speech via sensory-motor networks. Lip-reading also activated a much wider network in the superior temporal lobe than the sensory-motor task, possibly reflecting a more direct cross-sensory integration network.

In face-to-face conversations, we are sensitive not only to acoustic cues in the speech signal but also to the visual cues present in a speaker’s face. Lip-reading, or speech-reading, is the ability of humans to understand speech by observing the lip and mouth gestures of the speaker. Visual cues provide linguistic information beyond analysis of facial expressions and can facilitate speech perception when an auditory signal occurs in a noisy environment or is degraded by noise (Callan, Jones, Munhall, Callan, Kroos & Vatikiotis-Bateson, 2003; Dodd, 1977; Sumby & Pollack, 1954). Further, when audio and visual speech information are mismatched (hearing “ba” while watching someone articulate “ka”), this can induce a perceptual illusion, the McGurk effect (McGurk & MacDonald, 1976), whereby the acoustic speech information is incorrectly perceived. It is quite clear, therefore, that auditory and visual speech information interact in perception.

Recent neuroimaging work investigating the neural correlates of visual speech perception implicate auditory and language-related regions in the superior temporal lobe including the superior temporal sulcus (STS), and in some reports, primary auditory cortex (Calvert, Bullmore, Brammer, Campbell, Williams, McGuire, Woodruff, Iversen & David, 1997; Calvert & Campbell, 2003; Pekola, Ojanen, Autti, Jaaskelainen, Mottonen, Tarkiainen &

Correspondence: Gregory Hickok, Department of Cognitive Sciences, University of California at Irvine, Irvine, CA 92697-3800. Phone: 949-824-1409. Fax: 949-824-2307. E-mail: greg.hickok@uci.edu.

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.

Sams, 2005). While the STS is activated in virtually all studies, primary auditory areas are less reliably reported (Bernstein, Auer, Moore, Ponton, Don & Singh, 2002; Paulesu, Perani, Blasi, Silani, Borghese, Giovanni, Sensolo & Fazio, 2003). In most studies, speechreading also elicits activity of left lateralized or bilateral inferior frontal gyrus (IFG) and premotor cortex (Calvert & Campbell, 2003; Fridriksson, Moss, Davis, Baylis, Bonilha & Rorden, 2008; MacSweeney, Amaro, Calvert, Campbell, David, McGuire, Williams, Woll & Brammer, 2000; Paulesu, et al., 2003; Skipper, Nusbaum & Small, 2005; Skipper, van Wassenhove, Nusbaum, & Small, 2007), regions classically implicated in speech production.

A major conclusion coming from the work on the neural basis of audiovisual speech perception is that visual speech has its influence on the perception of heard speech via multisensory integration, and that the posterior STS is a critical region in this respect. For example, Calvert, Campbell and Brammer (2000) showed that a region in the posterior STS showed supra-additive responses to audio-visual speech ($AV > A \text{ alone} + V \text{ alone}$), which these authors considered to be a signature of multisensory integration (Calvert et al. 2000), although this claim is controversial (Hocking & Price, 2008). This view, that visual speech modulates auditory perception of speech via multisensory integration in the STS, aligns well with evidence from studies in both human and non-human primates that implicates the posterior STS in cross-sensory integration (Beauchamp et al., 2004; Barraclough et al., 2005; Noesselt et al., 2007; Stein and Stanford, 2008). A recent study recording single unit and local field potentials in monkeys while they perceived audiovisual monkey vocalizations provided direct evidence for the influence of STS activity on responses in auditory cortex, at least in monkeys, which supports this hypothesis (Ghazanfar, Chandrasekaran & Logothetis, 2008; Hikosaka, Iwai, Saito & Tanaka, 1988).

Although it seems clear that multisensory integration in the STS is a major contributor to audiovisual interactions in speech perception, it may not be the only source as neuroimaging evidence suggest a role for frontal motor-speech related areas (Paulesu et al., 2003, Skipper, et al. 2007). Behavioral evidence for this possibility comes from a study that found that a McGurk-like effect can be induced, not only by *viewing* incongruent speech gestures, but by the listener's own incongruent speech gestures (Sams et al., 2005). Listeners silently articulated speech sounds that were either congruent or incongruent with the syllables they were listening to. The incongruent condition led to significantly more misperceptions of the heard speech (32% correct) than the congruent condition (95% correct) suggesting that motor representations of speech can influence sensory perception of speech sounds. It has been suggested that the source of this influence is via efferent copies of motor commands that are transmitted to auditory regions, and that this process may form a kind of predictive (forward model) mechanism that modulates the analysis of sensory input (Sams et al., 2005; Skipper, et al. 2007; Poeppel et al., 2008). This purely motor effect, however, appears to be substantially weaker than visually-induced misperceptions, as the Sams et al. study found that when subjects viewed an audio-visual mismatch (the standard McGurk effect), performance dropped to 6% correct. Viewing one's own incongruent articulations in a mirror produced intermediate results (17% correct), showing that the addition of visual information associated with the same self-articulation resulted in additional performance decrement, which in turn suggests that visual information has an added influence beyond efferent motor copies.

Imaging evidence is consistent with the idea that some of the interaction between auditory and visual speech may be mediated by the motor system in that motor-speech related frontal structures, such as portions of Broca's area, typically activate during visual speech perception (see above). It is relevant that while both the posterior STS and Broca's area are activated during visual speech perception, their response properties are different under some conditions. For example, Miller and D'Esposito (Miller & D'Esposito, 2005) have shown that the posterior STS responds more to audiovisual speech that is perceived as fused than audiovisual speech

that is desynchronized and perceived as unfused. Broca's region showed the reverse pattern, responding more to unfused audiovisual speech, and with a later activity peak. Although it is not entirely clear how to interpret the details of these response patterns, it does suggest that pSTS and Broca's area are performing different kinds of computations on audiovisual speech stimuli.

Given the evidence reviewed above, we hypothesize the existence of two routes by which visual speech can influence auditory perception of speech sounds. One is via direct sensory-sensory integration in which visual speech information is integrated with auditory speech information in the STS via projections from sensory input systems (Calvert, et al., 2000; Ghazanfar, Chandrasekaran & Logothetis, 2008). The other route is via the motor system. Following previous authors (Paulesu et al., 1993; Poeppel, et al. 2008; Sams et al., 2005; Skipper et al. 2007), we hypothesize that visual speech gestures activate motor networks associated with articulating the visually perceived gestures, which in turn send efferent copies to sensory cortices via sensory-motor integration circuits, thus exerting an influence on perception. The behavioral evidence reviewed above suggests that this sensory-motor route is the weaker of the two.

The goal of the present study was to test one prediction of this hypothesis, namely that the perception of visual speech should activate networks known to be involved in sensory-motor integration for speech. A number of studies have investigated sensory-motor integration circuits for speech (Buchsbaum et al., 2001; Hickok et al., 2003; Buchsbaum et al., 2005a; Buchsbaum et al., 2005b). A network of brain regions has been identified that have both sensory and motor response properties in the speech domain, including portions of the STS, a region at the posterior most aspect of the Sylvian fissure at the parietal-temporal boundary (Spt), a portion of Broca's area (BA44 in particular), and a more dorsal premotor site in the frontal lobe.

The present study used standard paradigms for identifying sensory-motor activations for speech and for identifying cortical areas responsive to visual speech to determine if these two tasks activate partially overlapping networks. Such a finding would be consistent with the proposal that visual speech can influence heard speech via a sensory-motor mechanism.

Methods

Subjects

Twenty-six participants (10 females) between 18 and 44 years of age were recruited from the University of California, Irvine (UCI) community and received monetary compensation for their time. The volunteers were right-handed, native English speakers with normal or corrected-to-normal vision, no known history of neurological disease, and no other contraindications for MRI. Informed consent was obtained from each participant prior to participation in the study in accordance with guidelines from UCI Institutional Review Board which approved this study. Two subjects were omitted from data analysis due to excessive head motion and one subject was omitted for failing to follow task instructions.

Materials & Procedure

The data reported in this study were part of a larger experiment aimed at mapping responses to a range of sensory stimuli, including melodic sequences, noise bursts, auditory speech, and visual speech. Here we focus only on the speech conditions. The auditory speech stimuli were 3-second jabberwocky sentences (e.g., *It is the glandour in my nedderop*) in which content words replaced with nonsense words, as used in previous experiments (Hickok, Buchsbaum, Humphries & Muftuler, 2003). Visual speech stimuli were silent video clips of a male face

articulating six visually distinguishable CV syllables (ba, tha, va, bi, thi, vi). Following Calvert et al (1997), we also presented video clips of six lower-face non-speech gestures that were used as a control to isolate speech-reading activations. The non-speech gestures were: partial opening of the mouth with leftward deviation, opening of mouth with rightward deviation, opening of mouth with lip protrusion, tongue protrusion, lower lip biting, and lip retraction.

Subjects were randomly presented with 15-second blocks of music, speech, noise bursts or videos in equal ratios. Subjects were instructed to monitor for “oddball” stimuli throughout the study, where oddball was defined relative to the type of stimuli that constituted a given block. Oddball stimuli for the speech trials were a change in speaker voice from male to female, and for videos, a change in the actor from male to female. Subjects pressed a button each time an oddball was detected. There were 3 oddball trials within each session and in total this comprised approximately 13% of the experiment. These trials were modeled as a regressor of no interest and excluded from the results.

On some trials, subjects were simultaneously presented with 3-seconds of jabberwocky sentences and a picture of either an ear or lips which stayed on screen for 15 seconds. If a picture of lips was presented, subjects were instructed to rehearse the jabberwocky sentence until the lips went off screen. If a picture of an ear was presented, they were instructed to simply pay attention to the 3-second jabberwocky sentence and “rest” during the remainder of the 15-second block. The listen-rehearse condition has been shown to drive activity in sensory-motor regions of the posterior planum temporale (Spt) (Hickok, Buchsbaum, Humphries & Muftuler, 2003), and the listen-rest condition served as a control for the effects of acoustic stimulation alone. Thus “sensory-motor” activations are defined as regions that respond both during the perception of speech (a sensory response) and during covert rehearsal of speech (a “motor” response). Previous studies have confirmed that area Spt activates during more conventional motor-speech tasks such as picture naming (Graves, Grabowski, Mahta, & Gordon, 2007; Okada, Smith, Humphries & Hickok, 2003).

The experiment started with a short exposure session to familiarize subjects with all of the different experimental stimuli. Subjects were scanned during the exposure session to ensure they could comfortably hear the stimuli through the scanner noise, and to acclimatize them to the fMRI situation. This was followed by 5 experimental sessions (runs). Each experimental session contained an equal number of trials (blocks) of each condition. Each trial was 15 seconds in length and a single scanning session was approximately 6 minutes long. Auditory stimuli were presented through MR compatible headset and stimulus delivery and timing were controlled using Cogent software (http://www.vislab.ucl.ac.uk/cogent_2000.php) implemented in Matlab 6 (Mathworks, Inc, USA).

Scanning parameters

MR images were obtained in a Philips Achieva 3T (Philips Medical Systems, Andover, MA) fitted with an 8 channel RF receiver head coil, at the high field scanning facility at the University of California, Irvine. We first collected a total of 608 EPI volumes over 5 sessions using Fast Echo EPI (sense reduction factor=2.0, matrix=112×112mm, TR=3.0s, TE=25ms, flip angle=70°, size=1.95×1.95×2mm). After the functional scans, a high resolution anatomical image was acquired in axial plane (matrix=256×256mm, TR=8ms, TE=3.7ms, flip angle=8°, size=1×1×1mm).

Data Analysis

Preprocessing of the data was performed using Statistical Parametric Mapping (SPM5; Wellcome Department of Imaging Neuroscience, London, UK; www.fil.ion.ucl.ac.uk/spm) implemented in Matlab7 (Mathworks, Inc, USA). First, motion correction was performed by

creating a mean image from all of the volumes in the experiment and then realigning all volumes to that mean image using a 6-parameter rigid-body model. Images were smoothed with an isotropic 4 mm full width half maximum (FWHM) Gaussian kernel. To facilitate group analysis, images were spatially normalized to a standard template based on the Montreal Neurological Institute (MNI) reference brain.

First level analysis was performed on each subject using AFNI software (Cox, 1996). Regression analysis was performed to find parameter estimates that best explained variability in the data. Each predictor variable representing the time course of stimulus presentation was convolved with a gamma variate function and entered into the general linear model. The five regressors used in the present analysis are the following: oddballs (regressor of no interest), jabberwocky sentences, listen-rehearse, listen-rest, speech videos, face gesture videos. To test specific hypotheses, linear contrasts were performed and T-statistics computed at each voxel to identify regions significantly activated in the lip-reading task and the sensory-motor task. To identify cortical regions of enhanced activity associated with lip-reading, viewing speech was contrasted with viewing lower-face gestures ($p < .001$). To find regions specifically activated in the sensory-motor task, listen-rehearse was contrasted with listen-rest ($p < .001$). Using these two masks (each contrast thresholded at $p < .001$), we identified regions of direct overlap in the sensory-motor and lip-reading tasks. Group-level analysis was then performed on the linear contrasts of the parameter estimates, treating each subject as a random effect ($p < .001$).

Results

In all 23 subjects, the lip-reading task (relative to facial gurning) activated a large portion of superior temporal gyrus bilaterally including portions of the STS as well as more dorsal regions in the anterior and posterior supratemporal plane, the left posterior inferior frontal gyrus, and left and right dorsal premotor cortex (see Figure 1A and Table 1). The sensory-motor task (listen→rehearse - listen→rest) also activated a more focal region of the STG including portions of the STS and a region in the posterior planum temporale (Spt), left posterior parietal cortex, bilateral inferior frontal gyrus, and left dorsal premotor cortex (see Figure 1A and Table 1). Regions of overlap between the two contrasts (conjunction of sensory-motor task and lip-reading task) were found in the left inferior frontal gyrus, left dorsal premotor cortex, left posterior STS, left anterior STG (see Figure 1A and Table 2), and at a slightly relaxed threshold ($p = .005$), in the left posterior planum temporale (Spt) (Figure 1B). As the variable anatomy in the planum temporale region often reduces power in group averaged analyses, a relaxed threshold is justified in this situation (Westbury et al., 1999).¹ This constellation of regions comprises the network previously identified and hypothesized to support sensory-motor integration for speech.

Discussion

Consistent with previous studies, lip-reading activated bilateral STG, the posterior supratemporal plane, and speech-related regions in the frontal cortex. Comparison of the activation pattern for lip-reading with that of a sensory-motor speech task (listening to and then covertly repeating heard speech) showed that a portion of the lip-reading activation directly overlapped with sensory-motor integration regions for speech. Specifically, in a conjunction analysis, the lip-reading and sensory-motor tasks (relative to their respective controls) jointly activated the posterior IFG, dorsal premotor cortex, Spt in the planum temporale region, and focal portions of the STS, all in the left hemisphere. This shows that visual speech activates

¹Analysis was performed in individual subjects and we found that the overlap was present at $p < .001$. The overlap is therefore not an artifact of group averaging.

sensory-motor integration networks and therefore is consistent with the view that visual speech can influence heard speech via this system, perhaps in the form of efferent copies of motor predictions (Paulesu et al., 1993; Poeppel, et al. 2008; Sams et al., 2005; Skipper et al. 2007).

We also found substantial activation within the superior temporal lobe for lip-reading that did not overlap activation found in the sensory-motor task. This finding is consistent with the proposal that the influence of visual speech on heard speech is not restricted to sensory-motor interactions, but may also influence heard speech via cross-sensory integration (Calvert, et al. 2000; Ghazanfar, Chandrasekaran & Logothetis, 2008).

We suggest that visual speech activates both a sensory-sensory integration network in the STS and a sensory-motor network including frontal motor-speech regions, area Spt, and focal regions of the STS. We further propose that both of these networks provide independent sources of constraint on the analysis of acoustic speech input, although the cross-sensory system appears to be the most influential.

References

- Barracough NE, Xiao D, Baker CI, Oram MW, Perrett DI. Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *Journal of Cognitive Neuroscience* 2005;17:377–391. [PubMed: 15813999]
- Beauchamp MS, Lee KE, Argall BD, Martin A. Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* 2004;41:809–823. [PubMed: 15003179]
- Beauchamp M, Argall BD, Bodurka J, Duyn JH, Martin A. Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nature Neuroscience* 2004;7:1190–1193.
- Bernstein LE, Auer ET, Moore JK, Ponton CW, Don M, Singh M. Visual speech perception without primary auditory cortex activation. *Neuroreport* 2002;13:311–315. [PubMed: 11930129]
- Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET. Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex* 2000;10:512–528. [PubMed: 10847601]
- Buchsbaum B, Hickok G, Humphries C. Role of Left Posterior Superior Temporal Gyrus in Phonological Processing for Speech Perception and Production. *Cognitive Science* 2001;25:663–678.
- Buchsbaum BR, Olsen RK, Koch P, Berman KF. Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron* 2005a;48:687–697. [PubMed: 16301183]
- Buchsbaum BR, Olsen RK, Koch PF, Kohn P, Kippenhan JS, Berman KF. Reading, hearing, and the planum temporale. *Neuroimage* 2005b;24:444–454. [PubMed: 15627586]
- Callan DE, Jones JA, Munhall K, Callan AM, Kroos CK, Vatikiotis-Bateson E. Neural processes underlying perceptual enhancement by visual speech gestures. *Neuroreport* 2003;14:2213–2218. [PubMed: 14625450]
- Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SCR, McGuire PK, Woodruff PWR, Iversen SD, David AS. Activation of auditory cortex during silent lip-reading. *Science* 1997;276:593–596. [PubMed: 9110978]
- Calvert GA, Campbell R. Reading speech from still and moving faces: The neural substrates of visible speech. *Journal of Cognitive Neuroscience* 2003;15:57–70. [PubMed: 12590843]
- 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. The processing of audio-visual speech: empirical and neural bases. *Philosophical Transactions of The Royal Society B* 2008;363:1001–1010.
- Dodd B. The role of vision in the perception of speech. *Perception* 1977;6:31–40. [PubMed: 840618]
- Dodd B. Lip-reading in Infants: Attention to speech presented in- and out- of synchrony. *Cognitive Psychology* 1979;11:478–484. [PubMed: 487747]

- Fridriksson F, Moss J, Davis B, Baylis GC, Bonilha L, Rorden C. Motor speech perception modulates the cortical language areas. *NeuroImage* 2008;41:605–613. [PubMed: 18396063]
- Galaburda A, Sanides F. Cytoarchitectonic organization of the human auditory cortex. *Journal of Comparative Neurology* 1980;190:597–610. [PubMed: 6771305]
- Ghazanfar AA, Chandrasekaran C, Logothetis NK. Interactions between the superior temporal sulcus and auditory cortex mediate dynamic face/voice integration in rhesus monkeys. *The Journal of Neuroscience* 2008;28:4457–4469. [PubMed: 18434524]
- Graves WW, Grabowski TJ, Mahta S, Gordon JK. A neural signature of phonological access: distinguishing the effects of word frequency from familiarity and length in overt picture naming. *Journal of Cognitive Neuroscience* 2007;19:617–631. [PubMed: 17381253]
- Griffiths TD, Warren JD. The planum temporale as a computational hub. *Trends in Neuroscience* 2002;25:348–353.
- Grossman ED, Battelli L, Pascual-Leone A. Repetitive TMS over posterior STS disrupts perception of biological motion. *Vision Research* 2005;45:2847–2853. [PubMed: 16039692]
- Grossman ED, Blake R. Brain areas active during visual perception of biological motion. *Neuron* 2002;35:1167–1175. [PubMed: 12354405]
- Hickok G, Buchsbaum B, Humphries C, Muftuler T. Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience* 2003;15:673–682. [PubMed: 12965041]
- Hikosaka K, Iwai E, Saito H, Tanaka K. Polysensory properties of neurons in the anterior bank of caudal superior temporal sulcus of the macaque monkey. *Journal of Neurophysiology* 1988;60:1615–1637. [PubMed: 2462027]
- Hocking J, Price CJ. The role of posterior superior temporal sulcus in audiovisual processing. *Cerebral Cortex* 2008;18:2439–2449. [PubMed: 18281303]
- MacSweeney M, Amaro E, Calvert GA, Campbell R, David AS, McGuire P, Williams SCR, Woll B, Brammer MJ. Silent speechreading in the absence of scanner noise: an fMRI study. *Neuroreport* 2000;11:1729–1733. [PubMed: 10852233]
- McGurk H, MacDonald J. Hearing lips and seeing voices. *Nature* 1976;264:746–748. [PubMed: 1012311]
- 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]
- Noesselt T, Rieger JW, Schoenfeld MA, Kanowski M, Hinrichs H, Heinze HJ, Driver J. Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *Journal of Neuroscience* 2007;27:11431–11441. [PubMed: 17942738]
- Okada K, Smith KR, Humphries C, Hickok G. Word length modulates neural activity in auditory cortex during covert object naming. *Neuroreport* 2003;14:2323–2326. [PubMed: 14663184]
- Pa J, Hickok G. A parietal-temporal sensory-motor integration areas for the human vocal tract: evidence from an fMRI study of skilled musicians. *Neuropsychologia* 2007;46:362–368. [PubMed: 17709121]
- Paulesu E, Perani D, Blasi V, Silani G, Borghese NA, Giovanni UD, Sensolo S, Fazio F. A functional-anatomical model for lip-reading. *Journal of Neurophysiology* 2003;90:2005–2013. [PubMed: 12750414]
- Pekkola J, Ojanen V, Autti T, Jaaskelainen IP, Mottonen R, Sams M. Attention to visual speech gestures enhances hemodynamic activity in the left planum temporale. *Human Brain Mapping* 2006;27:471–477. [PubMed: 16161166]
- Pekkola J, Ojanen V, Autti T, Jaaskelainen IP, Mottonen R, Tarkiainen A, Sams M. Primary auditory cortex activation by visual speech: an fMRI study. *Neuroreport* 2005;16:125–128. [PubMed: 15671860]
- Poeppel D, Idsardi WJ, van Wassenhove V. Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society Biological Sciences* 2008;363:1071–1086. [PubMed: 17890189]
- Sams M, Mottonen R, Sihvonen T. Seeing and hearing others and oneself talk. *Cognitive Brain Research* 2005;23:429–435. [PubMed: 15820649]
- Selzer B, Pandya DN. Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. *Brain Research* 1978;149:1–24. [PubMed: 418850]

- Skipper JI, Nusbaum HC, Small SL. Listening to talking faces: motor cortical activation during speech perception. *NeuroImage* 2005;25:76–89. [PubMed: 15734345]
- Skipper JI, van Wassenhove V, Nusbaum HC, Small SL. Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex* 2007;17:2387–2399. [PubMed: 17218482]
- Smiley JF, Hackett TA, Ulbert I, Karmas G, Lakatos P, Javitt DC, Schroeder CE. Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. *The Journal of Comparative Neurology* 2007;502:894–923. [PubMed: 17447261]
- Stein BE, Stanford TR. Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews Neuroscience* 2008;9:255–266.
- Sumby WH, Pollack I. Visual contributions to speech intelligibility in noise. *Journal of Acoustic Society of America* 1954;26:212–215.
- Warren JE, Wise RJ, Warren JD. Sounds do-able: auditory-motor transformations and the posterior temporal plane. *Trends in Neurosciences* 2005;28:636–643. [PubMed: 16216346]
- Westbury CF, Zatorre RJ, Evans AC. Quantifying variability in the planum temporale: a probability map. *Cerebral Cortex* 1999;9:392–405. [PubMed: 10426418]

Figure 1.

A. Conjunction map showing regions of enhanced activity in the lip-reading task ($p < .001$), the sensory-motor task ($p < .001$), and areas of overlap. Blue=lip-reading, Green=Sensory-Motor, Red=Overlap.

B. Regions showing significant overlap for the lip-reading and sensory-motor tasks relative to their respective baseline controls ($p < .005$). Crosshair at posterior planum temporale (Spt). Blue=lip-reading, Green=Sensory-Motor, Red=Overlap.

Table 1
Talairach coordinates of cortical regions activated by the lip-reading task and the sensory-motor task in group analysis.

Region	Approximate Brodmann's Area	Peak Location (x, y, z)
Lipreading Task	L Inferior Frontal Gyrus	(-48 12 28)
	L Precentral Gyrus/Inferior Frontal Gyrus	(-58 12 4)
	L Precentral Gyrus	(-56 -6 46)
	L posterior Middle Temporal Gyrus/Superior Temporal Sulcus	(-56 -28 2)
	L Superior Temporal Gyrus	(-66 -14 8)
	R Middle Frontal Gyrus	(52 22 30)
	R Superior Temporal Gyrus	(66 -26 0)
	L Inferior Frontal Gyrus	(-44 30 6)
	L Precentral Gyrus	(-56 -6 44)
	L Superior Frontal Gyrus	(-32 50 34)
Sensory-Motor Task	L Superior Frontal Gyrus/Medial Frontal Gyrus	(-2 4 56)
	L Superior Temporal Gyrus/L Inferior Frontal Gyrus	(-58 12 2)
	L Ant/Mid Middle Temporal Gyrus/Superior Temporal Sulcus	(-64 -14 -6)
	L Posterior Superior Temporal Gyrus	(-64 -34 18)
	L Posterior Middle Temporal Gyrus/Superior Temporal Sulcus	(-62 -50 6)
	L Inferior Parietal Lobe	(-56 -42 28)
	L Inferior Parietal Lobe	(-56 -36 44)
	R/L Medial Frontal Gyrus	(0 58 38)
	R/L Medial Frontal Gyrus	(0 16 44)
	R Superior Temporal Gyrus/R Inferior Frontal Gyrus	(52 14 2)

Talairach coordinates of cortical regions showing overlap in the group analysis: conjunction of lip-reading task and sensory-motor tasks.

Table 2

Region	Approximate Brodmann's Area	Peak Location (x, y, z)
L Middle Frontal Gyrus/Precentral Gyrus	6	(-44 2 38)
L Medial Frontal Gyrus	6	(-2 0 58)
L Inferior Frontal Gyrus		(-36 32 -2)
L Posterior Superior Temporal Gyrus/Spt		(-62 -36 14)
L Ant/Mid Middle Temporal Gyrus/Superior Temporal Sulcus	21	(-62 -18 -4)
L Superior Temporal Gyrus	42/22	(-70 -22 6)
L Middle Temporal Gyrus	22	(-64 -32 4)
L Posterior Middle Temporal Gyrus/Superior Temporal Sulcus	22	(-62 -44 4)
R Superior Temporal Gyrus/Inferior Frontal Gyrus		(54 12 -2)
R Superior Temporal Gyrus	22	(58 8 -4)
R Ant/Mid Middle Temporal Gyrus/Superior Temporal Sulcus		(62 -16 -6)
R Posterior Middle Temporal Gyrus/Superior Temporal Sulcus	21	(64 -30 -4)