

Dynamic Changes in Superior Temporal Sulcus Connectivity during Perception of Noisy Audiovisual Speech

Audrey R. Nath and Michael S. Beauchamp

Department of Neurobiology and Anatomy, University of Texas Medical School at Houston, Houston, Texas 77030

Humans are remarkably adept at understanding speech, even when it is contaminated by noise. Multisensory integration may explain some of this ability: combining independent information from the auditory modality (vocalizations) and the visual modality (mouth movements) reduces noise and increases accuracy. Converging evidence suggests that the superior temporal sulcus (STS) is a critical brain area for multisensory integration, but little is known about its role in the perception of noisy speech. Behavioral studies have shown that perceptual judgments are weighted by the reliability of the sensory modality: more reliable modalities are weighted more strongly, even if the reliability changes rapidly. We hypothesized that changes in the functional connectivity of STS with auditory and visual cortex could provide a neural mechanism for perceptual reliability weighting. To test this idea, we performed five blood oxygenation level-dependent functional magnetic resonance imaging and behavioral experiments in 34 healthy subjects. We found increased functional connectivity between the STS and auditory cortex when the auditory modality was more reliable (less noisy) and increased functional connectivity between the STS and visual cortex when the visual modality was more reliable, even when the reliability changed rapidly during presentation of successive words. This finding matched the results of a behavioral experiment in which the perception of incongruent audiovisual syllables was biased toward the more reliable modality, even with rapidly changing reliability. Changes in STS functional connectivity may be an important neural mechanism underlying the perception of noisy speech.

Introduction

A key problem for both humans and computers is how to understand noisy speech, the most important method of human communication (Hauser, 1996; Kryter, 1996; Dupont and Luettnin, 2000). Multisensory integration is likely to be important for understanding noisy speech: because the information coming from the auditory and visual modalities is independent, combining across modalities reduces noise and allows for more accurate perception (Sumbly and Pollack, 1954; Stein and Meredith, 1993). However, little is known about the neural mechanisms for multisensory integration during perception of noisy speech.

In humans, the superior temporal sulcus (STS) integrates auditory and visual information about both speech and nonspeech stimuli (Calvert et al., 2000; Sekiyama et al., 2003; Wright et al., 2003; Beauchamp et al., 2004; Callan et al., 2004; Miller and D'Esposito, 2005; Stevenson and James, 2009; Werner and Noppeney, 2010). Anatomical studies in nonhuman primates have shown that the STS receives input from both auditory cortex and extrastriate visual cortex (Seltzer et al., 1996; Lewis and Van Essen, 2000). However, anatomical connections can vary in strength under different behavioral circumstances, a property that has been characterized as functional connectivity (McIntosh

and Gonzalez-Lima, 1994; Büchel and Friston, 2001; Horwitz et al., 2005; Stein et al., 2007; de Marco et al., 2009).

Behavioral studies have shown that perception of multisensory stimuli is reliability weighted: information from the more reliable modality is given a stronger weight (Ernst and Banks, 2002; Alais and Burr, 2004; Ma et al., 2009). We hypothesized that during perception of noisy audiovisual speech, the functional connectivity between the STS and sensory cortex would be modulated by the amount of noise present in each sensory modality. We predicted that this modulation would match the pattern observed behaviorally. More reliable (less noisy) sensory stimuli would result in stronger connectivity between that sensory cortex and STS, while less reliable (more noisy) sensory stimuli would result in weaker connectivity between that sensory cortex and STS.

To test this hypothesis, we performed behavioral and functional magnetic resonance imaging (fMRI) experiments in which subjects were presented with more reliable and less reliable audiovisual speech stimuli consisting of syllables and words. Independent functional localizers were used to identify STS, auditory cortex, and extrastriate visual cortex regions of interest in each subject. Then, in separate scan series, functional connectivity was measured between STS and sensory cortex. A behavioral experiment was conducted to verify that our noisy audiovisual speech produced the same reliability weighting observed in previous behavioral studies (Ernst and Banks, 2002; Alais and Burr, 2004; Ma et al., 2009). The first fMRI experiment measured STS functional connectivity during perception of blocks of words that were reliable or unreliable in the auditory and visual modalities. However, behavioral studies have demonstrated that reliability weighting

Received Sept. 15, 2010; revised, accepted Nov. 16, 2010.

Correspondence should be addressed to Dr. Michael S. Beauchamp, 6431 Fannin Street, Suite G.550, Houston, TX 77030. E-mail: Michael.S.Beauchamp@uth.tmc.edu.

This research was supported by National Science Foundation Grant 642532 and National Institutes of Health Grants R01NS065395, TL1RR024147, and S10RR019186. We thank nine anonymous reviewers and the editors for their helpful comments and Vips Patel for assistance with MR data collection.

DOI:10.1523/JNEUROSCI.4853-10.2011

Copyright © 2011 the authors 0270-6474/11/310001-01\$15.00/0

AQ: A

En1

AQ: V

Table 1. Stimuli and tasks across the five experiments

	fMRI experimental design	Auditory-only	Visual-only	Auditory-reliable	Visual-reliable	Task
Functional localizer	Blocked	Undegraded words (C)	Undegraded words (C)	n/a	n/a	Passive
Experiment 1	Blocked	n/a	n/a	Words (C)	Words (C)	Passive
Experiment 2	Event-related	n/a	n/a	Words (C)	Words (C)	Passive
Experiment 3	Event-related	n/a	n/a	n/a	Undegraded (C + I) Midblur (C + I) High blur (C + I)	C versus I C versus I C versus I
Experiment 4	Event-related	n/a	n/a	Syllables (C)	Syllables (C)	Attn-A: "ja" versus "ma" Attn-V: eyes open versus closed
Experiment 5	n/a (behavioral)	n/a	n/a	Syllables (C + I)	Syllables (C + I)	"Ma" versus "na"

C, Congruent; I, incongruent; Attn-A, auditory attention; Attn-V, visual attention; n/a, not applicable.

can change dynamically from trial to trial, and can vary parametrically (Ernst and Banks, 2002). Therefore, in the second and third fMRI experiments, STS functional connectivity was measured while stimulus reliability was manipulated at one of two levels (second experiment) or three levels (third experiment) and changed dynamically in successive word presentations using a rapid event-related design. Behavioral studies have also demonstrated that reliability weighting occurs even in the presence of sustained attention to one modality (Helbig and Ernst, 2008). Therefore, in the fourth fMRI experiment, functional connectivity was measured while subjects' attention was directed toward or away from the auditory or visual modalities. Across experiments, a consistent pattern of reliability-weighted STS functional connectivity was observed, suggesting that functional connectivity is a plausible candidate for the neural mechanism of behavioral reliability weighting.

Materials and Methods

Subjects and stimuli

Thirty-four healthy right-handed subjects participated in one behavioral and four fMRI experiments (13 female, mean age 27.6; 10 subjects in experiment 1, 10 in experiment 2, 6 in experiment 3, 6 in experiment 4, 10 in experiment 5). The subjects provided informed written consent under an experimental protocol approved by the Committee for the Protection of Human Subjects of the University of Texas Health Science Center at Houston.

Table 1 summarizes the stimulus conditions for each experiment, which consisted of single-word or single-syllable speech presented in the auditory modality, the visual modality, or both. A digital video system was used to record a female speaker saying 200 single-syllable words from the MRC Psycholinguistic Database with Brown verbal frequency of 20–200, imageability rating >100, age of acquisition <7 years, and Kucera–Francis written frequency >80 (Wilson, 1988). The duration of the words ranged from 0.5 to 0.7 s. The total length of each video clip ranged from 1.1 to 1.8 s to start and end each video with the speaker in a neutral, mouth-closed position and to include all mouth movements from mouth opening to closing.

Multisensory audiovisual speech consisted of both the auditory and visual recordings presented synchronously (for sample stimuli, see Fig. 1A,B). Unisensory visual speech consisted of the video portion of the speech, followed by a still image of a scrambled face presented for 50 ms to minimize afterimages. Unisensory auditory speech consisted of the audio portion of the speech and a visual display consisting of white fixation crosshairs. The baseline condition consisted of only the fixation crosshairs; the crosshairs were presented at the same position as the mouth during visual speech to minimize eye movements.

Reliable (non-noisy) speech consisted of the original recordings without any degradation. Unreliable (noisy) speech was created by independently degrading the auditory and visual components of the recordings. The auditory stimulus was degraded in Matlab (MathWorks) using a noise-vocoded filter to modulate noise within the temporal envelope of the stimulus in four separate frequency bands (0–800 Hz, 800–1500 Hz, 1500–2500 Hz, and 2500–4000 Hz), and the smoothing frequency, or

sampling rate of the temporal envelope of each frequency band of the noise-vocoded speech, was 300 Hz (Shannon et al., 1995). The visual stimulus was degraded by individually filtering each frame of the video in Matlab. First, the contrast of the frame was decreased by 70%, and then the frame was blurred with a Gaussian filter (the filter size depended on the stimulus condition; see below). Both of these methods for degrading speech decrease intelligibility: vocoding auditory speech stimuli decreases recognition scores (Shannon et al., 1995), and decreasing the spatial resolution of visual speech stimuli decreases word identification (Munhall et al., 2004).

General fMRI methods

At the beginning of each scanning session, two T1-weighted MP-RAGE anatomical MRI scans were collected at 3 tesla using an eight-channel head gradient coil; the anatomical scans were aligned to each other and averaged to provide maximum gray-white contrast. Then, a cortical surface model was created with FreeSurfer (Dale et al., 1999; Fischl et al., 1999) to allow visualization and region-of-interest creation with SUMA (Argall et al., 2006). T2*-weighted images for fMRI were collected using gradient-echo echo-planar imaging (TR = 2015 ms, TE = 30 ms, flip angle = 90°) with in-plane resolution of 2.75 × 2.75 mm. Thirty-three 3 mm axial slices were collected, resulting in whole-brain coverage in most subjects. Each functional scan series consisted of 154 brain volumes. The first four volumes, collected before equilibrium magnetization was reached, were discarded resulting in 150 usable volumes. Ear defender (acoustic earmuff) type pneumatic headphones were used to present auditory stimuli within the scanner, ensuring stimulus audibility even without the use of a sparse sampling paradigm. Visual stimuli were projected onto a screen using an LCD projector and viewed through a mirror attached to the head coil. Behavioral responses were collected using a fiber-optic button response pad (Current Designs). MR-compatible eye tracking (Applied Science Laboratories) was used in all fMRI experiments to ensure alertness and visual fixation.

fMRI data analysis was performed using the freely available software packages "R" (Ihaka and Gentleman, 1996) and Analysis of Functional NeuroImages software (AFNI) (Cox, 1996). Appendices A–C contain the complete list of commands used to analyze data in experiments 1 and 2 to facilitate replication of our analyses. Corrections for voxelwise multiple comparisons were performed using the false discovery rate procedure (Genovese et al., 2002) and reported as "q" values. Data were analyzed within each subject, and then group analyses were performed by combining data across subjects using a random-effects model.

Functional data were aligned to the average anatomical dataset and motion corrected for each voxel in each subject using a local Pearson correlation (Saad et al., 2009). All analysis was performed in all voxels in each subject in the context of the generalized linear model using a maximum-likelihood approach using the AFNI function 3dDeconvolve.

For block designs (localizer and experiment 1), a single regressor was created for each stimulus type by convolving the stimulus timing with a canonical gamma-variate estimate of the hemodynamic response function. For rapid event-related designs (experiments 2, 3, and 4), the amplitude of the hemodynamic response for each individual auditory-reliable and visual-reliable stimulus presentation was estimated. A separate regressor was created for each individual stimulus using the -stim_times_IM mode of 3dDeconvolve. Movement covariates and

T1

F1

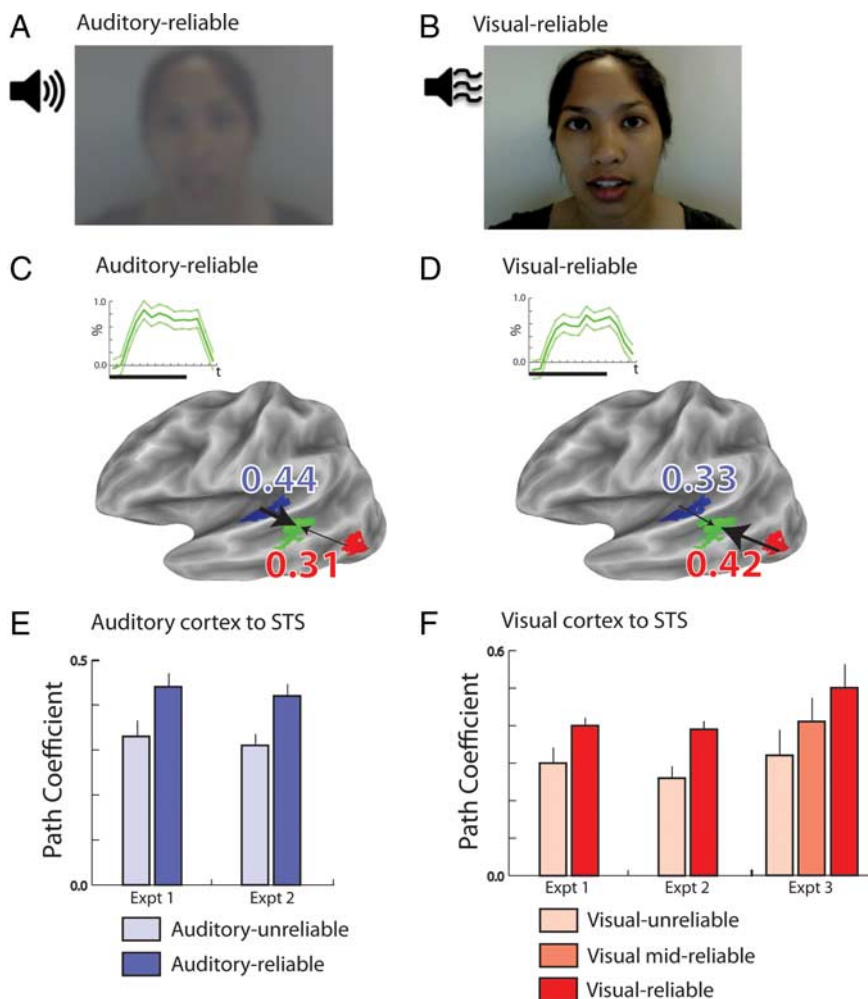


Figure 1. Speech stimuli and functional connectivity. **A**, Auditory-reliable speech: undegraded auditory speech (loudspeaker icon) with degraded visual speech (single video frame shown). **B**, Visual-reliable speech: undegraded visual speech with degraded auditory speech. **C**, STS BOLD response and functional connectivity during auditory-reliable speech in a representative subject. The inset graph shows the BOLD percentage signal change to blocks of speech (thick green line shows mean response; thin green lines show SEM; thick black bar shows 20 s duration of each block). The green region on the cortical surface model shows the location of the STS ROI. The blue number is the path coefficient between the STS ROI and the auditory cortex ROI (shown in blue); the red number is the path coefficient between the STS ROI and the visual cortex ROI (shown in red). Lateral view of the partially inflated left hemisphere, dark gray shows sulcal depths, light gray shows gyral crowns. **D**, STS BOLD response and functional connectivity during visual-reliable speech, same subject as in **C**. **E**, STS–auditory cortex functional connectivity in experiments 1, 2, 4A (attention to auditory modality), and 4B (attention to visual modality). Light blue bars, Path coefficients during auditory-unreliable speech. Dark blue bars, Path coefficients during auditory-reliable speech. **F**, STS–visual cortex functional connectivity in experiments 1–4.

Table 2. ROIs from fMRI localizer

ROI	Size (mm ³)	Talairach coordinates (mm)		
		x	y	z
Auditory	3108 ± 969	−45.4 ± 4.1	−17.5 ± 3.6	6.1 ± 2.7
Visual	3111 ± 1068	−42.4 ± 3.6	−66.5 ± 4.9	3.1 ± 3.9
STS	3611 ± 1210	−48.8 ± 2.9	−46 ± 6.4	9.7 ± 3.8

Average size and location of individual auditory, visual, and STS ROIs created from functional localizers across all subjects (mean ± SD). Sample subject ROIs are shown in Figure 1, C and D.

baseline drifts (as second-order polynomials, one per scan series) were modeled as regressors of no interest. All regressors were fit separately to each voxel.

For volumetric group analysis, each subject’s average anatomical dataset was normalized to the N27 reference anatomical volume (Mazziotta et al., 2001) using the auto_tlrc function in AFNI.

fMRI functional localizer and regions of interest

A functional localizer scan series was used to identify regions of interest (ROIs) important for speech processing: auditory cortex, visual cortex, and STS. Each ROI was identified separately in each subject (for ROI sizes and locations, see Table 2). The ROIs were then applied to independent reliability-weighting data collected in separate scan series, preventing bias (Kriegeskorte et al., 2009). ROIs were created in both left and right hemispheres, although our analysis focused on the left hemisphere because the left hemisphere is usually dominant for language (Branch et al., 1964; Ellmore et al., 2010).

The functional localizer scan series contained 10 blocks (five unisensory auditory and five unisensory visual in random order) of duration 20 s with 10 s of fixation baseline between each block. Each block contained ten 2 s trials, each containing one undegraded word. All ROIs were created on the cortical surface using SUMA for maximum anatomical precision. The auditory cortex ROI was defined using the contrast of auditory speech versus baseline to find active voxels within Heschl’s gyrus, demarcated by the superior temporal sulcus at the lateral/inferior boundary, the first temporal sulcus at the anterior boundary, the transverse temporal sulcus at the posterior boundary, and the medial termination of Heschl’s gyrus at the medial/superior boundary (Patterson and Johnsrude, 2008; Upadhyay et al., 2008). The visual cortex ROI was defined using the contrast of visual speech versus baseline to find active voxels along the inferior temporal sulcus (ITS) or its posterior continuation near areas LO and MT (Dumoulin et al., 2000) within extrastriate lateral occipital cortex, a brain region critical for processing moving and biological stimuli that includes the middle temporal visual area and the extrastriate body area (Tootell et al., 1995; Beauchamp et al., 1997, 2002, 2003; Downing et al., 2001; Pelphey et al., 2005). The STS ROI was defined using a conjunction analysis to find all voxels that responded to both auditory and visual speech significantly greater than baseline in the anatomically defined posterior STS ($q < 0.05$ for each modality) (Beauchamp, 2005a; Beauchamp et al., 2008).

Structural equation modeling

To test the hypothesis that connection weights would be different for auditory-reliable and visual-reliable speech, a structural equation model was constructed and tested for each stimulus condition in each subject in each experiment. Path coefficients from the models were compared across subjects using an ANOVA.

The model consisted of the three ROIs (auditory cortex, visual cortex, and STS) in the left hemisphere with unidirectional connections between auditory cortex and STS and between visual cortex and STS (for ROIs in one subject, see Fig. 1C,D). Models were also tested consisting of bidirectional connections between the left-hemisphere ROIs and unidirectional connections between right-hemisphere ROIs. In addition, we performed a whole-brain connectivity analysis on the experiment 1 data to determine whether any other regions outside the auditory and visual cortex ROIs showed condition-dependent changes in connection strength with STS, similar to a psychophysiological interaction (Friston et al., 1997).

AQ:B, T2

AQ: C

AQ: U

The whole-brain connectivity analysis was first performed in each subject and then averaged across subjects using an ANOVA.

Experiment 1: fMRI block design

fMRI experimental design. To minimize the number of stimulus conditions in experiments 1 and 2, reliable stimuli in one modality were paired with unreliable stimuli in the other modality, resulting in only two conditions: auditory-reliable (auditory-reliable + visual-unreliable) and visual-reliable (visual-reliable + auditory-unreliable). Unreliable videos were created by blurring with a 30 pixel diameter Gaussian filter. Three reliability scan series were presented to each subject. Scan series contained five blocks of auditory-reliable and five blocks of visual-reliable congruent words. Each 20 s block contained ten 2 s trials, with one different word per trial (10 words total per block). Each video ranged in length from 1.1 to 1.8 s with fixation crosshairs occupying the remainder of each 2 s trial. Each block contained a different set of randomly selected words. There were 10 s of fixation baseline between each block, for a total scan series time of 5 min.

fMRI data analysis. As shown in Figure 1, stimulus blocks evoked a robust blood oxygenation level-dependent (BOLD) signal consisting of a typical square-wave-like response. A generalized linear model analysis was used to calculate the average amplitude of response to each stimulus type in each ROI. However, analysis of fMRI time series containing these large block onsets and offsets produces a very high correlation between ROIs that obscures differences in functional connectivity (Büchel and Friston, 1997). Therefore, for the connectivity analysis in each subject, these block onset and offset responses (defined as the best fit of the block-design regressors) were subtracted from the time series in each voxel, as were the best-fit regressors of no interest for that voxel, using the AFNI function 3dSynthesize (see Appendix A for the details of all functions used in the analysis). Because a distinct set of areas is active during baseline conditions in the so-called default-mode network (Raichle et al., 2001), our connectivity analysis only examined the 10 volumes per block collected during word presentation (150 time points for each condition, consisting of 15 blocks of each stimulus type times 10 time points per block). The residual time series for each condition was averaged across all voxels in each ROI to create two time series for each ROI (one per condition). The residual time series were used to calculate the correlation matrix and path coefficients for each condition using the AFNI functions Iddot and Idsem (Chen et al., 2007). Two-way ANOVAs were performed in MATLAB with condition (auditory-reliable or visual-reliable) and sensory cortex (auditory or visual) as fixed factors, subject as a random factor, and amplitude of response and path coefficient as the dependent measures.

Whole-brain connectivity: psychophysiological interactions. The ROI analysis determined whether the connectivity between auditory cortex, visual cortex, and STS differed across conditions. To determine whether any other brain regions showed differing connectivity with the STS, we performed a whole-brain connectivity analysis similar to a psychophysiological interaction (Friston et al., 1997). The time series for each condition was created consisting of the concatenated residual time series for the auditory-reliable and visual-reliable stimulation blocks (see Appendix B for details). A generalized linear model analysis was performed on this dataset using three regressors of interest. The first regressor (physiological factor) consisted of the residual time series from the STS ROI. The second regressor (psychological factor) consisted of 150 values of +1 (corresponding to the 150 visual-reliable time points) and 150 values of -1 (corresponding to the 150 auditory-reliable time points). The third regressor (the psychophysiological factor) consisted of the first regressor multiplied by the second regressor. Voxels with a significant amount of variance accounted for by the third regressor show a different slope relating activity in the STS to activity in the target region as a function of condition. The *t*-statistic of the psychophysiological factor and the *t*-statistic of the full model from each individual subject were averaged across subjects using a voxelwise ANOVA to find all voxels that showed a significant response to a threshold of all word stimuli ($q < 0.0095$) and a significant difference in correlation with the STS across conditions ($q < 0.05$).

Experiment 2: fMRI rapid event-related design

Two reliability scan series were presented to each subject, each containing 60 auditory-reliable word trials (2 s each), 60 visual-reliable word trials (2 s each), and 30 null trials (2 s of fixation baseline) presented pseudorandomly in optimal rapid event-related order (Dale, 1999). The amplitude of the hemodynamic response was estimated for each individual word stimulus and averaged within each ROI to produce a vector of 60 auditory-reliable word amplitudes and 60 visual-reliable word amplitudes (see Appendix C for details). These amplitudes were used to calculate the correlation matrix and path coefficients for each condition in each subject using the AFNI functions Iddot and Idsem. The path coefficients were then entered into the group ANOVA.

Experiment 3: fMRI rapid event-related parametric design

To create a parametric design in experiment 3, the auditory unreliable words from experiments 1 and 2 were paired with visual words that were either the same word (congruent; 50% of trials) or a different word (incongruent). Subjects made a two-alternative forced choice about each stimulus using a button press (congruent vs incongruent). The visual words were reliable (unblurred) or blurred with one of three Gaussian filter widths: 5 pixels, 15 pixels, or 30 pixels. Three reliability scan series were presented to each subject. Each scan series contained 30 trials of each of the four stimulus types (120 total) and 30 trials of fixation baseline in optimized order. The behavioral and fMRI data from the 5 and 15 pixel stimuli did not differ significantly, so they were collapsed for the final analysis, resulting in three levels: reliable visual (unblurred); unreliable, blurred visual (30-pixel-width blur, the same as used in experiments 1 and 2); and intermediate midblur (5- and 15-pixel-width blur). The three conditions were then analyzed as in experiment 2.

Experiment 4: attention experiment

Although the event-related designs of experiments 2 and 3 eliminated the effect of sustained attention over the course of a block, subjects could still have reallocated attention on a trial-by-trial basis. Therefore, in experiment 4, we manipulated the behavioral task to explicitly direct subjects' attention to either the auditory or visual modality during presentation of auditory-reliable and visual-reliable syllables. A rapid event-related design was used in which auditory-reliable and visual-reliable congruent syllable stimuli were randomly intermixed. Subjects performed a behavioral two-alternative forced-choice task that directed their attention to either modality.

The stimuli consisted of two instances of syllable "ja" and two instances of "ma," one with the speaker's eyes open and one with the speaker's eyes closed. For the auditory attention task, subjects discriminated the syllables ("ja" vs "ma"). For the visual attention task, subjects discriminated the visual appearance of the speaker (eyes open vs eyes closed).

Our choice of different auditory and visual tasks was motivated by several competing desires. First, we wished subjects to maintain attention to the auditory or visual speech, to ensure a high level of activity in STS. This ruled out an orthogonal discrimination task, such as detection of brightness changes at fixation or auditory beeps. Second, we wished to maintain ethological validity by using tasks that were not too different from cognitive processes that might occur during normal speech processing. We did not wish to use tasks that were so difficult that they might drive a high level of activity, obscuring the brain activity related to speech perception. This ruled out complex semantic decision tasks. Third, we wished to make information from the opposing modality uninformative for the task. There was absolutely no information about eyes open or eyes closed in the auditory stimulus, so subjects' optimal strategy would be to ignore the auditory modality, presumably enhancing any effect of attention and reducing any effect of reliability weighting (if reliability weighting depended on voluntary attention to the auditory modality). In contrast, if we had asked subject to perform a visual "ja" from "ma" discrimination, they could have "cheated" by using auditory information. To prevent cheating, we would have had to introduce incongruent auditory-visual stimuli, which would in itself cause changes in brain activity, since the STS is very sensitive to auditory-visual congruence (van Atteveldt et al., 2010).

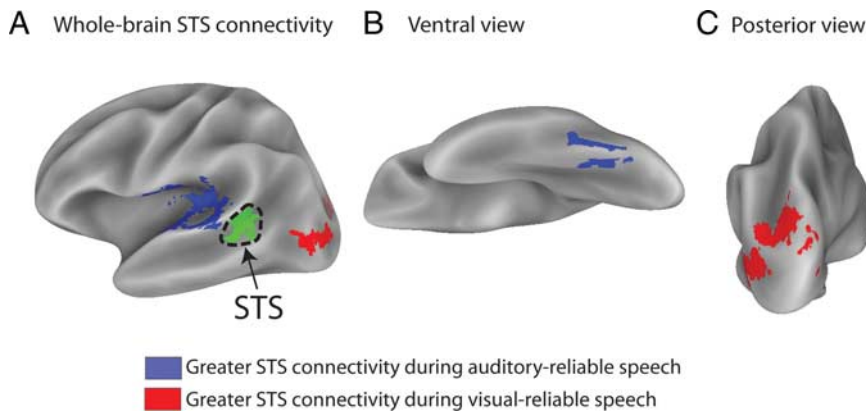


Figure 2. Whole-brain connectivity group map. **A**, Whole-brain connectivity analysis showing regions with differential connectivity with STS during auditory-reliable and visual-reliable words. Shown is a group map from 10 subjects with STS seed region shown in green surrounded by dashed line. Blue areas showed greater connectivity with the STS during auditory-reliable speech, and red areas showed greater connectivity during visual-reliable speech. Shown is a lateral view of the partially inflated average cortical surface, left hemisphere. **B**, Ventral view of the left hemisphere showing a region near the location of the fusiform face area that showed stronger connections with the STS during auditory-reliable speech. **C**, Dorsal view of the left hemisphere showing a region of dorsal occipital cortex, near visual area V3A, with greater STS connectivity during visual-reliable speech.

Table 3. Whole-brain connectivity analysis

Interaction	Brain region	Size (mm ³)	Talairach coordinates		
			x	y	z
Auditory-reliable	L STG	1427	-63	-33	8
	L fusiform gyrus	210	-43	-67	-16
Visual-reliable	L LOC	202	-43	-79	2
	L V3a	142	-17	-93	12

Regions in the experiment 1 group dataset showing a positive interaction with STS during auditory-reliable blocks or visual-reliable blocks. The regions are illustrated in Figure 2. L, Left.

Each scan series contained 120 stimuli and 30 presentations of the baseline condition in a random order. Within the 120 stimuli in each scan series, there were equal numbers of auditory-reliable and visual-reliable stimuli, equal numbers of syllables corresponding to “ja” and “ma,” and equal numbers of stimuli with the speaker’s eyes open and closed. Subjects performed the auditory attention task for two scan series and the visual attention task for two scan series. Four separate structural equation models were constructed, one for each condition in each attentional state.

Experiment 5: behavioral experiment

A behavioral experiment was performed to make sure that our stimuli elicited the behavioral reliability weighting observed in previous studies. Ten healthy right-handed subjects (two female, mean age 24.8) participated in the behavioral study. To prevent ceiling effects, all stimuli were degraded more than the stimuli used in the fMRI experiments. The auditory-reliable stimuli consisted of auditory syllables noise-vocoded with a smoothing frequency of 1000 Hz and visual-unreliable syllables blurred with a 50 pixel diameter Gaussian filter, and the visual-reliable stimuli consisted of auditory-unreliable syllables noise-vocoded with a smoothing frequency of 300 Hz and visual syllables blurred with a 30 pixel diameter Gaussian filter. Congruency was manipulated in experiment 5 so that half the stimuli were congruent and half were incongruent. A 2 × 2 × 2 factorial design was used, with auditory syllable (“ma” vs “na”) as the first factor, visual syllable (“ma” vs “na”) as the second factor, and reliability (auditory-reliable vs visual-reliable) as the third factor. Auditory stimuli were delivered through headphones at ~70 dB, and visual stimuli were presented on a computer screen. Each of 10 subjects was presented with 80 stimuli (10 examples of each stimulus type) and made a two-alternative forced choice about their perception of each stimulus (“ma” vs “na”). Responses to incongruent stimuli (e.g., auditory “ma” + visual “na”) were analyzed with a within-subjects paired *t* test.

Results

The STS responded to both auditory-reliable and visual-reliable audiovisual speech with a robust BOLD response (Fig. 1C,D), but there was no difference in the amplitude of the response: 0.62% for auditory-reliable versus 0.69% for visual-reliable (*p* = 0.18, paired *t* test across subjects).

Although the amplitude of the BOLD response was similar between conditions, there was a large change in STS functional connectivity: the STS was more strongly connected to a sensory cortex when that sensory modality was reliable (Fig. 1E,F). STS–auditory cortex connectivity increased from 0.33 for auditory-unreliable to 0.44 for auditory-reliable (*p* = 0.02, paired *t* test) for the blocks of words presented in experiment 1 and from 0.31 to 0.42 (*p* = 0.003) for the single words presented in an event-related design in experiment 2. STS–visual cortex connectivity increased from 0.30 for visual-unreliable to 0.40 for visual-reliable (*p* = 0.04) in experiment 1 and from 0.26 to 0.39 (*p* = 0.0003) in experiment 2. In experiment 3, in which a parametric design was used with single words presented at three levels of visual reliability, STS–visual cortex connectivity increased from 0.32 to 0.42 to 0.50 with increasing levels of visual reliability (ANOVA with main effect of reliability: *F*_(1,5) = 17.9, *p* = 0.0005).

STS–visual cortex connectivity increased from 0.30 for visual-unreliable to 0.40 for visual-reliable (*p* = 0.04) in experiment 1 and from 0.26 to 0.39 (*p* = 0.0003) in experiment 2. In experiment 3, in which a parametric design was used with single words presented at three levels of visual reliability, STS–visual cortex connectivity increased from 0.32 to 0.42 to 0.50 with increasing levels of visual reliability (ANOVA with main effect of reliability: *F*_(1,5) = 17.9, *p* = 0.0005).

Whole-brain connectivity: psychophysiological interactions

Our initial analysis measured the connection strength between the STS and auditory and visual cortex ROIs created from independent functional localizers. To determine whether other brain areas also showed reliability-weighted connections, we performed a *post hoc* whole-brain connectivity analysis that searched for brain areas showing stimulus-dependent interactions with the STS (Fig. 2, Table 3).

Regions with greater STS connectivity during auditory-reliable speech were concentrated in and around auditory cortex, while regions with a greater STS connectivity during visual-reliable speech were concentrated in lateral occipital cortex. These regions largely corresponded to the auditory and visual ROIs generated from the localizer. Additional regions showing differential connectivity were found in the fusiform gyrus (greater STS connectivity during auditory-reliable speech) and dorsal occipital cortex, near visual area V3A (greater STS connectivity during visual-reliable speech). Notably, neither calcarine cortex (the location of V1) nor portions of Heschl’s gyrus (the location of primary auditory cortex) showed condition-dependent changes in connectivity.

BOLD amplitudes

The observed increases in functional connectivity with reliability could be driven by differential responses in sensory cortex. The amplitudes of the BOLD response in sensory cortex increased significantly for more reliable stimuli in experiments 1 and 2 [auditory cortex, experiment (Expt.) 1: 0.46% to 0.72%, *p* = 0.000003; Expt. 2: 0.18% to 0.36%, *p* = 0.0003; visual cortex: 0.38% to 0.51%, *p* = 0.001; 0.23% to 0.29%, *p* = 0.002] but not experiment 3 (0.32% to 0.34% to 0.37%, *F*_(1,5) = 2.1, *p* = 0.18). If

F2, T3

there was a simple relationship between STS connectivity and sensory cortex BOLD amplitude, we might expect to find a correlation across subjects between connectivity and amplitude. However, there was little correlation between the two in any experiment, with all p values >0.20 (auditory cortex, correlation between STS functional connectivity and BOLD amplitude, Expt. 1, $r = 0.14$, $p = 0.70$ for auditory-unreliable stimuli and $r = 0.44$, $p = 0.20$ for auditory-reliable stimuli, Expt. 2, $r = -0.15$, $p = 0.68$ and $r = -0.06$, $p = 0.87$; visual cortex, correlation between STS functional connectivity and BOLD amplitude, Expt. 1, $r = 0.16$, $p = 0.85$ for visual-unreliable stimuli and $r = -0.08$, $p = 0.66$ for visual-reliable stimuli, Expt. 2, $r = -0.41$, $p = 0.25$ and $r = -0.10$, $p = 0.78$; Expt. 3, $r = -0.21$, $p = 0.69$ for visual-unreliable, $r = 0.18$, $p = 0.73$ for visual mid-reliable, and $r = -0.42$, $p = 0.41$ for visual-reliable).

Attention experiment

Another possible cause of reliability-related changes in connection weights could be top-down factors, especially attention (Büchel and Friston, 1997; Friston and Büchel, 2000; Macaluso et al., 2000; Mozolic et al., 2008). In experiment 4, the behavioral task was manipulated to direct subjects' attention to either the auditory or visual modality during presentation of auditory-reliable or visual-reliable speech. Subjects accurately performed the instructed task (97% correct during auditory attention and 93% correct during visual attention), indicating that subjects attended to the correct modality. Despite the presence of directed attention, the changes in STS connectivity observed in experiments 1, 2, and 3 were replicated (STS–auditory cortex: 0.32 to 0.50, $p = 0.00003$; STS–visual cortex: 0.25 to 0.40, $p = 0.002$). An ANOVA demonstrated no interaction between attention and reliability ($F_{(1,5)} = 0.11$, $p = 0.74$).

Additional analyses

Structural equation models depend on initial assumptions about connections within the network. When we modified the model to include bidirectional connections or right hemisphere regions of interest, we observed similar, but weaker, patterns of reliability weighting. For bidirectional connections, STS–auditory cortex connectivity increased from 0.55 to 0.60 ($p = 0.10$) in experiment 1 and from 0.45 to 0.53 ($p = 0.01$) in experiment 2, and STS–visual cortex connectivity increased from 0.55 to 0.59 ($p = 0.13$) in experiment 1, 0.45 to 0.50 ($p = 0.01$) in experiment 2, and 0.51 to 0.62 to 0.68 ($F_{(1,5)} = 39.1$, $p = 0.00002$) in experiment 3. For right hemisphere regions of interest, STS–auditory cortex connectivity increased from 0.39 to 0.47 ($p = 0.15$) in experiment 1 and from 0.40 to 0.50 ($p = 0.007$) in experiment 2, and STS–visual cortex connectivity increased from 0.28 to 0.32 ($p = 0.42$) in experiment 1, 0.20 to 0.26 ($p = 0.03$) in experiment 2, and 0.45 to 0.40 to 0.53 ($F_{(1,5)} = 1.9$, $p = 0.21$) in experiment 3.

Behavioral experiment

To replicate previous studies demonstrating that the perception of audiovisual speech is driven by the more reliable sensory modality, in experiment 5 we created incongruent stimuli whose reliability was altered using the techniques of experiments 1–4. When subjects were presented with incongruent stimuli that were reliable in one modality and unreliable in the other modality, they were more likely to classify the stimulus as the syllable presented in the reliable modality ($p = 0.0001$, paired t test).

Discussion

The most important result of these experiments was the surprising finding that the functional connectivity of the STS changes

dramatically during perception of noisy speech, depending on the reliability of the auditory and visual modalities. The changes in connectivity were striking: the dominant modality, defined as the sensory modality with the strongest input to STS, was determined by reliability. The changes in functional connectivity were rapid, happening within 2 s during the rapid event-related experiments (experiments 2, 3, and 4) in which auditory-reliable and visual-reliable speech was randomly intermixed. Consistent results were observed across a variety of stimuli and behavioral tasks in four experiments, suggesting that the phenomenon of reliability-weighted STS connectivity is not dependent on a particular stimulus or task.

An obvious candidate to produce rapid and large changes in functional connectivity between STS and sensory cortex is the activity within sensory cortex itself. Consistent with this idea, we observed increases in the BOLD amplitude of response in sensory cortex for more reliable stimuli. This is consistent with previous fMRI studies in which auditory speech degraded using a noise-vocoded filter (as used in our study) resulted in reduced activity in auditory cortex (Scott et al., 2000; Narain et al., 2003; Giraud et al., 2004; Davis et al., 2005; Obleser et al., 2007) and low-contrast images (such as the blurred videos in our study) resulted in reduced activity in visual cortex (Callan et al., 2004; Olman et al., 2004; Park et al., 2008; Stevenson et al., 2009). While the changes in BOLD amplitude in sensory cortex in our study can be parsimoniously explained as reflecting low-level stimulus properties, we did not observe a significant correlation between STS connectivity and sensory cortex BOLD amplitude. One possible explanation is that STS neurons are driven by only a subset of neurons within sensory cortex, with a normalization that divides the strongest response in the input population by the pooled background activity (Ghose, 2009; Lee and Maunsell, 2009; Reynolds and Heeger, 2009). If we could more accurately measure the activity of this subset of neurons, perhaps using MR adaptation techniques (van Atteveldt et al., 2010), a stronger relationship between activity and connectivity might be observed.

These results can also be interpreted in light of predictive coding models of cortical function (Kersten et al., 2004). The BOLD signal in sensory cortex is higher when a correct inference (hit) is made about auditory or visual stimuli than during misses of identical stimuli or false alarms (Hesselmann et al., 2010), suggesting that the BOLD signal in sensory cortex could be a measure of the brain's confidence about the perceptual hypothesis represented by neurons in that sensory cortex. In this model, the STS could use this confidence measure to adjust its own predictive model of the multisensory environment by adjusting its connection weights with sensory cortex.

The experimental findings are consistent with the idea that the STS is a critical brain area for auditory–visual multisensory integration (Beauchamp, 2005b). In macaque STS, a region known as STP (superior temporal polysensory) or TPO (temporo-parietal-occipital) receives projections from auditory and visual association cortex (Seltzer et al., 1996; Lewis and Van Essen, 2000) and contains single neurons that show enhanced responses to auditory and visual communication signals (Dahl et al., 2009). For brevity, we have referred to the human homolog of this region as “STS” while noting that the STS also contains other functionally and anatomically heterogeneous regions (Beauchamp, 2005b; Van Essen, 2005; Hein and Knight, 2008). During speech perception, the auditory cortex processes spectral and temporal information from the auditory vocalization, extrastriate visual cortex processes cues from lip movements, and the STS integrates the auditory and visual information (Binder et al., 1997; Price, 2000;

Scott and Johnsrude, 2003; Belin et al., 2004; Hickok and Poeppel, 2007; Zatorre, 2007; Bernstein et al., 2008, 2010; Campbell, 2008; Poeppel et al., 2008). Interrupting activity in the STS reduces the McGurk effect, an illusion that depends on auditory–visual interactions (Beauchamp et al., 2010b), supporting a role for the STS in auditory–visual multisensory integration during speech (Scott and Johnsrude, 2003; Miller and D’Esposito, 2005; Campanella and Belin, 2007).

A number of behavioral studies show that the presence of information from both the auditory and visual modalities aids the perception of noisy speech (Sumbly and Pollack, 1954; Hauser, 1996; Kryter, 1996; Grant and Seitz, 2000; MacDonald et al., 2000; Shahin and Miller, 2009). The perception of multisensory stimuli is reliability weighted: information from the more reliable modality is given a stronger weight (experiment 5) (Ernst and Banks, 2002; Alais and Burr, 2004). The connectivity differences we observed followed this same pattern, with more reliable sensory stimuli producing stronger connectivity between that sensory cortex and STS. If both modalities provide equivalent amounts of information, then the neural signals representing those modalities should be weighted equally. In contrast, if one modality provides poor quality information, it should receive less weighting by multisensory areas such as the STS. Therefore, it seems plausible that reliability-weighted functional connectivity between STS and sensory cortex could be the neural substrate for the reliability weighting observed behaviorally. This brain mechanism for understanding noisy speech may also be applicable to machine-learning strategies for computer-based speech recognition (Dupont and Luettin, 2000; Girin et al., 2001).

For the initial analysis, a structural equation model was selected in which auditory and visual cortex provide unidirectional projections to STS. However, there are both top-down and bottom-up connections throughout the cortical processing hierarchy (Felleman and Van Essen, 1991; Murray et al., 2002; de la Mothe et al., 2006; Winer, 2006; van Atteveldt et al., 2009). When incorporating bidirectional connections into the structural equation model, we also observed reliability weighting. Models of the right hemisphere showed weaker reliability weighting, perhaps reflecting the less important role of the right hemisphere in speech perception (Wolmetz et al., 2011).

Even without the constraints of the functional localizers (used to identify sensory cortex regions of interest), the whole-brain connectivity analysis demonstrated reliability-weighted connectivity changes between STS and auditory and extrastriate visual cortex. Interestingly, the whole-brain analysis also suggested that connectivity between core regions of auditory cortex and primary visual cortex were not reliability weighted. This may reflect the anatomical finding that STS receives strong visual input from extrastriate visual areas such as MT, but not V1, and that STS receives stronger input from auditory association areas than from core areas of auditory cortex (Seltzer and Pandya, 1994; Lewis and Van Essen, 2000; Smiley et al., 2007). A provocative finding in our dataset was the increased connection weight between STS and regions of ventral temporal cortex (near the fusiform face area) during auditory-reliable stimulation. If this region forms a node in the network for person identification (Kanwisher and Yovel, 2006; von Kriegstein et al., 2008), and auditory information is especially useful for person identification when visual information is degraded, then it would be behaviorally advantageous to increase connection weights between the fusiform face area and STS.

Behavioral studies have shown that reliability weighting occurs even if subjects are forced to attend to one modality, suggest-

ing that reliability weighting is independent of modality-specific attention (Helbig and Ernst, 2008). Consistent with this finding, in experiment 4 we found that reliability-weighted connection changes persisted even if subjects’ attention was directed to one modality or the other. Because we observed the same pattern of connectivity changes in experiments with passive word presentation (experiments 1 and 2) and with three different behavioral tasks (congruence detection in experiment 3; visual discrimination and auditory discrimination in experiment 4), attention or behavioral context is unlikely to be the sole explanation of our results.

We define “noisy” and “unreliable” stimuli operationally, as stimuli with reduced intelligibility (perceptual accuracy). Vo-coded auditory speech (Davis et al., 2005; Dahan and Mead, 2010) and blurred visual speech (Thomas and Jordan, 2002; Gordon and Allen, 2009) have been used in many previous studies to reduce intelligibility. A recent study of visual–tactile integration used dynamic visual white noise, instead of blurring, to degrade the visual modality and found similar reliability-weighted changes in functional connectivity (Beauchamp et al., 2010a), suggesting that the precise method used to degrade stimuli is unlikely to explain our results.

Changes in functional connectivity have been observed in other studies of multisensory integration (Hampson et al., 2002; Horwitz and Braun, 2004; Fu et al., 2006; Patel et al., 2006; Gruber et al., 2007; Kreifelts et al., 2007; Noesselt et al., 2007; Obleser et al., 2007; Noppeney et al., 2008), with stronger weights most often observed in conditions in which multisensory stimuli result in behavioral improvements. Kreifelts et al. (2007) found that connection weights from sensory cortex to multisensory areas increased in strength during multisensory stimulation compared with unisensory stimulation. Noesselt et al. (2007) observed greater functional coupling between sensory cortex and STS when audiovisual stimuli were temporally congruent than when they were not. In our study, the sensory cortex processing the more reliable modality had a stronger connection with STS. Together with these previous fMRI studies, our results suggest that increased functional coupling could be a general mechanism for promoting multisensory integration. Under situations in which multisensory integration occurs, as codified by Stein and Meredith’s laws of multisensory integration (Stein and Meredith, 1993), connection strengths between sensory cortex and multisensory areas are expected to be strong.

Appendix A: Commands for Analyzing Experiment 1 Data

The following are the commands used to analyze a single-subject dataset for experiment 1.

First, we aligned the two T1 anatomical scans (“\${ec}” refers to the subject’s experiment code, the code for each patient used to preserve anonymity).

```
3dAllineate -base 3dsag_t1_2.nii -source 3dsag_t1_1.nii -prefix
${ec}anatr1_1RegTo2 -verb -warp shift_rotate -cost mi -auto-
mask -1Dfile ${ec}anatr2toanatr1
```

Then, we averaged the two aligned anatomical scans into one dataset.

```
3dmerge -gnzmean -nscale -prefix ${ec}anavg 3dsag_t1_2.nii
${ec}anatr1_1RegTo2+orig
```

We created skull-stripped anatomy to which the EPI could be aligned.

```
3dSkullStrip -input ${ec}anavg+orig -prefix ${ec}anavgSS
```

The skull-stripped anatomy was then normalized to the N27 reference anatomical volume.

AQ: F

AQ: G

AQ: E

AQ: H

```
@auto_tlrc -base TT_N27+tlrc -no_ss -input {$Sec}anatavgSS+orig
adwarp -apar {$Sec}anatavgSS_at+tlrc -dpar {$Sec}anatavg+orig1ddot
```

The four fMRI scan series were concatenated to one file (1–3 refer to block design reliability runs, and 4 refers to the localizer run).

```
3dTcat -prefix {$Sec}rall_fmri_1.nii_fmri_2.nii_fmri_3.nii_fmri_4.nii
```

The next steps were for motion correction and distortion correction.

```
3dresample -master {$Sec}rall+orig -dxyz 1.0 0.938 0.938 -inset {$Sec}anatavgSS+orig -prefix {$Sec}anatavgSScrop
align_epi_anat.py -epi2anat -anat {$Sec}anatavgSS+orig -anat_has_skull no -epi {$Sec}rall+orig -epi_base mean
```

The aligned EPI data were smoothed using a $3 \times 3 \times 3$ mm FWHM Gaussian kernel.

```
3dmerge -doall -1blur_rms 3 -prefix {$Sec}Albl {$Sec}rall_al+orig
A mask of the EPI dataset was created.
```

```
3dAutomask -dilate 1 -prefix {$Sec}maskAlbl {$Sec}Albl+orig
fMRI activation maps were created through a deconvolution analysis.
```

```
3dDeconvolve -fout -tout -full_first -polort a -concat runs.txt -cbucket {$Sec}mr_bucket
```

```
-input {$Sec}Albl+orig -num_stimts 10 -nfirst 0 -jobs 2
```

```
-mask {$Sec}maskAlbl+orig
```

```
-stim_times 1 ARblock.txt 'BLOCK(20,1)' -stim_label 1 ARblock
```

```
-stim_times 2 VRblock.txt 'BLOCK(20,1)' -stim_label 2 VRblock
```

```
-stim_times 3 Ablock.txt 'BLOCK(20,1)' -stim_label 3 Ablock
```

```
-stim_times 4 Vblock.txt 'BLOCK(20,1)' -stim_label 4 Vblock
```

```
-stim_file 5 {$Sec}rall_vr_al_motion.1D'[0]' -stim_base 5
```

```
-stim_file 6 {$Sec}rall_vr_al_motion.1D'[1]' -stim_base 6
```

```
-stim_file 7 {$Sec}rall_vr_al_motion.1D'[2]' -stim_base 7
```

```
-stim_file 8 {$Sec}rall_vr_al_motion.1D'[3]' -stim_base 8
```

```
-stim_file 9 {$Sec}rall_vr_al_motion.1D'[4]' -stim_base 9
```

```
-stim_file 10 {$Sec}rall_vr_al_motion.1D'[5]' -stim_base 10
```

```
-prefix {$Sec}mr
```

The next steps describe the structural equation modeling of the block design data. First, we used 3dSynthesize to find the best fit of the block-design regressors.

```
3dSynthesize -cbucket {$Sec}mr_bucket+orig -matrix {$Sec}mr.xmat.1D -select all -prefix {$Sec}mr_all
```

Then, the best fits of the block-design regressors were subtracted from the aligned and smoothed EPI dataset to create a residual dataset.

```
3dcalc -a {$Sec}Albl+orig -b {$Sec}mr_all+orig -prefix {$Sec}mr_noall -expr "a-b"
```

The trials in the residual dataset corresponding to the auditory-reliable (AR) and visual-reliable (VR) blocks were extracted into separate files.

```
3dTcat -prefix {$Sec}AR_noall {$Sec}mr_noall+orig'[300..309,315..324,375..384,405..414,435..444,450..459,465..474,480..489,495..504,570..579,615..624,645..654,690..699,705..714,735..744]'
```

```
3dTcat -prefix {$Sec}VR_noall {$Sec}mr_noall+orig'[330..339,345..354,360..369,390..399,420..429,510..519,525..534,540..549,555..564,585..594,600..609,630..639,660..669,675..684,720..729]'
```

These residual time series were used to calculate the correlation matrix for each condition. The rows (left to right) and columns (top to bottom) refer to the auditory ROI, visual ROI, and STS ROI.

```
3dROIstats -quiet -mask {$Sec}_L_ROI+orig {$Sec}AR_noall+orig > AR.1D
```

```
1ddot -terse AR.1D
```

```
1.00000 0.58155 0.58876
```

```
0.58155 1.00000 0.51082
```

```
0.58876 0.51082 1.00000
```

```
3dROIstats -quiet -mask {$Sec}_L_ROI+orig {$Sec}VR_noall+orig > VR.1D
```

```
1ddot -terse VR.1D
```

```
1.00000 0.58355 0.60122
```

```
0.58355 1.00000 0.67784
```

```
0.60122 0.67784 1.00000
```

The output from 1ddot was then entered into the afni function 1dsem, which prompted the user for the number and names of the ROIs, the correlation matrix (as shown above), and the parameters of the structural equation model. For the unidirectional model, we specified unidirectional connections from auditory and visual cortex to STS; for the bidirectional model, we specified bidirectional connections. The output of the SEM from each subject was then entered into an ANOVA in Matlab for group analysis.

Appendix B: Commands for Whole-Brain Connectivity Analysis

The following are the commands used to perform the whole-brain connectivity analysis (PPI) for experiment 1. First, the commands in Appendix A were executed, followed by these additional commands.

The auditory-reliable and visual-reliable residual time courses were concatenated together.

```
3dTcat -prefix {$Sec}RelWt_noall {$Sec}AR_noall+orig {$Sec}VR_noall+orig
```

A generalized linear model analysis was performed on this dataset using three regressors of interest: the residual time series from the STS ROI, 150 values of +1 (150 auditory-reliable time points) and 150 values of -1 (150 visual-reliable time points), and the first regressor multiplied by the second regressor (all regressors created in Microsoft Excel and saved as text files).

```
3dDeconvolve -fout -tout -full_first -polort a
```

```
-input {$Sec}RelWt_noall+orig -num_stimts 9 -nfirst 0 -jobs 2
```

```
-mask {$Sec}maskAlbl+orig
```

```
-stim_file 1 STSnoall.txt -stim_label 1 STSnoall
```

```
-stim_file 2 condition.txt -stim_label 2 condition
```

```
-stim_file 3 PPI_noall.txt -stim_label 3 PPI_noall
```

```
-stim_file 4 {$Sec}rall_vr_al_motion.1D'[0]' -stim_base 4
```

```
-stim_file 5 {$Sec}rall_vr_al_motion.1D'[1]' -stim_base 5
```

```
-stim_file 6 {$Sec}rall_vr_al_motion.1D'[2]' -stim_base 6
```

```
-stim_file 7 {$Sec}rall_vr_al_motion.1D'[3]' -stim_base 7
```

```
-stim_file 8 {$Sec}rall_vr_al_motion.1D'[4]' -stim_base 8
```

```
-stim_file 9 {$Sec}rall_vr_al_motion.1D'[5]' -stim_base 9
```

```
-prefix {$Sec}PPImr
```

The single-subject output from each subject was then entered into a group analysis with the following command:

```
3dANOVA2 -overwrite -type 3 -avelevs 2 -blevels 10
```

```
-dset 1 1 FPPPImr+tlrc'[8]' -dset 2 1 FPPPImr+tlrc'[10]'
```

```
-dset 1 2 FRPPPImr+tlrc'[8]' -dset 2 2 FRPPPImr+tlrc'[10]'
```

```
-dset 1 3 FTPPPImr+tlrc'[8]' -dset 2 3 FTPPPImr+tlrc'[10]'
```

```
-dset 1 4 FVPPPImr+tlrc'[8]' -dset 2 4 FVPPPImr+tlrc'[10]'
```

```
-dset 1 5 FXPPPImr+tlrc'[8]' -dset 2 5 FXPPPImr+tlrc'[10]'
```

```
-dset 1 6 GFPPPImr_AA1+tlrc'[8]' -dset 2 6 GFPPPImr_AA1+tlrc'[10]'
```

```
-dset 1 7 GGPPPImr_AA1+tlrc'[8]' -dset 2 7 GGPPPImr_AA1+tlrc'[10]'
```

```
-dset 1 8 GIPPImr+tlrc'[8]' -dset 2 8 GIPPImr+tlrc'[10]'
```

```
-dset 1 9 GKPPPImr+tlrc'[8]' -dset 2 9 GKPPPImr+tlrc'[10]'
```



```
-dset 1 10 GMPPImr+tlrc'[8]' -dset 2 10 GMPPImr+tlrc'[10]'
-fa Stimuli -amean 1 PPI_Tstat -amean 2 AVFullFstat
-bucket 3dANOVA_PPI
```

Appendix C: Commands for Analyzing Experiment 2 Data

The following are the commands used to analyze a single-subject dataset for experiment 2.

First, we aligned the two T1 anatomical scans (“ec” refers to the subject’s experiment code).

```
3dAllineate -base 3dsag_t1_2.nii -source 3dsag_t1_1.nii -prefix
${ec}anatr1_1RegTo2 -verb -warp shift_rotate -cost mi -auto-
mask -1Dfile ${ec}anatr2toanatr1
```

Then, we averaged the two aligned anatomical scans into one dataset.

```
3dmerge -gnzmean -nscale -prefix ${ec}anavg 3dsag_t1_2.nii
${ec}anatr1_1RegTo2+orig
```

We created skull-stripped anatomy to which the EPI could be aligned.

```
3dSkullStrip -input ${ec}anavg+orig -prefix ${ec}anavgSS
```

The skull-stripped anatomy was then normalized to the N27 reference anatomical volume.

```
@auto_tlrc -base TT_N27+tlrc -no_ss -input ${ec}anavgSS+
orig
```

```
adwarp -apar ${ec}anavgSS_at+tlrc -dpar ${ec}anavg+
orig1ddot
```

The four fMRI scan series are concatenated to one file (1–3 refer to block design reliability runs, and 4 refers to the localizer run).

```
3dTcat -prefix ${ec}rall_fmri_1.nii_fmri_2.nii_fmri_3.nii_fmri_
4.nii
```

The next steps are for motion correction and distortion correction.

```
3dresample -master ${ec}rall+orig -dxyz 1.0 0.938 0.938 -inset
${ec}anavgSS+orig -prefix ${ec}anavgSScrop
align_epi_anat.py -epi2anat -anat ${ec}anavgSS+orig -anat_
has_skull no -epi ${ec}rall+orig -epi_base mean
```

The aligned EPI data are smoothed using a $3 \times 3 \times 3$ mm FWHM Gaussian kernel.

```
3dmerge -doall -1blur_rms 3 -prefix ${ec}Albl ${ec}rall_al+orig
A mask of the EPI dataset is created.
```

```
3dAutomask -dilate 1 -prefix ${ec}maskAlbl ${ec}Albl+orig
```

The amplitude of the hemodynamic response was estimated for each individual word stimulus.

```
3dDeconvolve -fout -tout -full_first -polort a -concat runs.txt
-cbucket ${ec}RERmr_bucket
```

```
-input ${ec}Albl+orig -num_stimts 10 -nfirst 0 -jobs 2
```

```
-mask ${ec}maskAlbl+orig
```

```
-stim_times_IM 1 ARevent.txt 'BLOCK(2,1)' -stim_label 1
ARevent
```

```
-stim_times_IM 2 VRevent.txt 'BLOCK(2,1)' -stim_label 2
VRevent
```

```
-stim_times_IM 3 Ablock.txt 'BLOCK(20,1)' -stim_label 3
Ablock
```

```
-stim_times_IM 4 Vblock.txt 'BLOCK(20,1)' -stim_label 4
Vblock
```

```
-stim_file 5 ${ec}rall_vr_al_motion.1D'[0]' -stim_base 5
```

```
-stim_file 6 ${ec}rall_vr_al_motion.1D'[1]' -stim_base 6
```

```
-stim_file 7 ${ec}rall_vr_al_motion.1D'[2]' -stim_base 7
```

```
-stim_file 8 ${ec}rall_vr_al_motion.1D'[3]' -stim_base 8
```

```
-stim_file 9 ${ec}rall_vr_al_motion.1D'[4]' -stim_base 9
```

```
-stim_file 10 ${ec}rall_vr_al_motion.1D'[5]' -stim_base 10
```

```
-prefix ${ec}RERmr
```

The trials corresponding to the auditory-reliable (AR) and visual-reliable (VR) words are extracted into separate files.

```
3dbucket -prefix ${ec}AR '${ec}RERmr_bucket+orig[24-143]'
3dbucket -prefix ${ec}VR '${ec}RERmr_bucket+orig[144-263]'
```

These vectors of 60 auditory-reliable word amplitudes and 60 visual-reliable word amplitudes were used to calculate the correlation matrix for each condition. The rows (left to right) and columns (top to bottom) refer to the auditory ROI, visual ROI, and STS ROI.

```
3dROIstats -quiet -mask ${ec}_L_ROI_v1+orig ${ec}AR+
orig > AR.1D
```

```
1ddot -terse AR.1D
```

```
1.00000 0.69681 0.75244
```

```
0.69681 1.00000 0.70460
```

```
0.75244 0.70460 1.00000
```

```
3dROIstats -quiet -mask ${ec}_L_ROI_v1+orig ${ec}VR+
orig > VR.1D
```

```
1ddot -terse VR.1D
```

```
1.00000 0.43685 0.57950
```

```
0.43685 1.00000 0.67163
```

```
0.57950 0.67163 1.00000
```

The output from 1ddot was then entered into the afni function 1dsem, which prompted the user for the number and names of the ROIs, the correlation matrix (as shown above), and the parameters of the structural equation model. For the unidirectional model, we specified unidirectional connections from auditory and visual cortex to STS; for the bidirectional model, we specified bidirectional connections. The output of the SEM from each subject was then entered into an ANOVA in Matlab for group analysis.

References

Alais D, Burr D (2004) The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol* 14:257–262.

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

Beauchamp MS (2005a) Statistical criteria in fMRI studies of multisensory integration. *Neuroinformatics* 3:93–113.

Beauchamp MS (2005b) See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. *Curr Opin Neurobiol* 15:145–153.

Beauchamp MS, Cox RW, DeYoe EA (1997) Graded effects of spatial and featural attention on human area MT and associated motion processing areas. *J Neurophysiol* 78:516–520.

Beauchamp MS, Lee KE, Haxby JV, Martin A (2002) Parallel visual motion processing streams for manipulable objects and human movements. *Neuron* 34:149–159.

Beauchamp MS, Lee KE, Haxby JV, Martin A (2003) fMRI responses to video and point-light displays of moving humans and manipulable objects. *J Cogn Neurosci* 15:991–1001.

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.

Beauchamp MS, Yasar NE, Frye RE, Ro T (2008) Touch, sound and vision in human superior temporal sulcus. *Neuroimage* 41:1011–1020.

Beauchamp MS, Pasalar S, Ro T (2010a) Neural substrates of reliability-weighted visual-tactile multisensory integration. *Front Syst Neurosci* 4:25.

Beauchamp MS, Nath AR, Pasalar S (2010b) fMRI-guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. *J Neurosci* 30:2414–2417.

Belin P, Fecteau S, Bédard C (2004) Thinking the voice: neural correlates of voice perception. *Trends Cogn Sci* 8:129–135.

Bernstein LE, Lu ZL, Jiang J (2008) Quantified acoustic-optical speech signal incongruity identifies cortical sites of audiovisual speech processing. *Brain Res* 1242:172–184.

Bernstein LE, Jiang J, Pantazis D, Lu ZL, Joshi A (2010) Visual phonetic processing localized using speech and nonspeech face gestures in video

AQ: J

AQ: K

- and point-light displays. *Hum Brain Mapp*. Advance online publication. Retrieved ●. doi:10.1002/hbm.21139.
- Binder JR, Frost JA, Hammeke TA, Cox RW, Rao SM, Prieto T (1997) Human brain language areas identified by functional magnetic resonance imaging. *J Neurosci* 17:353–362.
- Branch C, Milner B, Rasmussen T (1964) Intracarotid sodium amytal for the lateralization of cerebral speech dominance; observations in 123 patients. *J Neurosurg* 21:399–405.
- Büchel C, Friston K (2001) Interactions among neuronal systems assessed with functional neuroimaging. *Rev Neurol (Paris)* 157:807–815.
- Büchel C, Friston KJ (1997) Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. *Cereb Cortex* 7:768–778.
- Callan DE, Jones JA, Munhall K, Kroos C, Callan AM, Vatikiotis-Bateson E (2004) Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. *J Cogn Neurosci* 16:805–816.
- Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol* 10:649–657.
- Campanella S, Belin P (2007) Integrating face and voice in person perception. *Trends Cogn Sci* 11:535–543.
- Campbell R (2008) The processing of audio-visual speech: empirical and neural bases. *Philos Trans R Soc Lond B Biol Sci* 363:1001–1010.
- AQ: L Chen G, Glen DR, Stein JL, Meyer-Lindenberg AS, Saad ZS, Cox RW (2007) Model validation and automated search in fMRI path analysis: a fast open-source tool for structural equation modeling. Paper presented at the Human Brain Mapping Conference, Chicago, IL, June.
- Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* 29:162–173.
- Dahan D, Mead RL (2010) Context-conditioned generalization in adaptation to distorted speech. *J Exp Psychol Hum Percept Perform* 36:704–728.
- Dahl CD, Logothetis NK, Kayser C (2009) Spatial organization of multisensory responses in temporal association cortex. *J Neurosci* 29:11924–11932.
- Dale AM (1999) Optimal experimental design for event-related fMRI. *Hum Brain Mapp* 8:109–114.
- Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9:179–194.
- Davis MH, Johnsrude IS, Hervais-Adelman A, Taylor K, McGettigan C (2005) Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. *J Exp Psychol Gen* 134:222–241.
- de la Mothe LA, Blumell S, Kajikawa Y, Hackett TA (2006) Cortical connections of the auditory cortex in marmoset monkeys: core and medial belt regions. *J Comp Neurol* 496:27–71.
- de Marco G, Vrignaud P, Destrieux C, de Marco D, Testelin S, Devauchelle B, Berquin P (2009) Principle of structural equation modeling for exploring functional interactivity within a putative network of interconnected brain areas. *Magn Reson Imaging* 27:1–12.
- Downing PE, Jiang Y, Shuman M, Kanwisher N (2001) A cortical area selective for visual processing of the human body. *Science* 293:2470–2473.
- Dumoulin SO, Bittar RG, Kabani NJ, Baker CL Jr, Le Goualher G, Bruce Pike G, Evans AC (2000) A new anatomical landmark for reliable identification of human area V5/MT: a quantitative analysis of sulcal patterning. *Cereb Cortex* 10:454–463.
- Dupont S, Luetjens J (2000) Audio-visual speech modeling for continuous speech recognition. *IEEE Trans Multimedia* 2:141–151.
- Ellmore TM, Beauchamp MS, Breier JI, Slater JD, Kalamangalam GP, O'Neill TJ, Disano MA, Tandon N (2010) Temporal lobe white matter asymmetry and language laterality in epilepsy patients. *Neuroimage* 49:2033–2044.
- Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415:429–433.
- Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex* 1:1–47.
- Fischl B, Sereno MI, Dale AM (1999) Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage* 9:195–207.
- Friston KJ, Büchel C (2000) Attentional modulation of effective connectivity from V2 to V5/MT in humans. *Proc Natl Acad Sci U S A* 97:7591–7596.
- Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997) Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage* 6:218–229.
- Fu CHY, McIntosh AR, Kim J, Chau W, Bullmore ET, Williams SCR, Honey GD, McGuire PK (2006) Modulation of effective connectivity by cognitive demand in phonological verbal fluency. *Neuroimage* 30:266–271.
- Genovese CR, Lazar NA, Nichols T (2002) Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* 15:870–878.
- Ghose GM (2009) Attentional modulation of visual responses by flexible input gain. *J Neurophysiol* 101:2089–2106.
- Giraud AL, Kell C, Thierfelder C, Sterzer P, Russ MO, Preibisch C, Kleinschmidt A (2004) Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. *Cereb Cortex* 14:247–255.
- Girin L, Schwartz J-L, Feng G (2001) Audio-visual enhancement of speech in noise. *J Acoust Soc Am* 109:3007–3020.
- Gordon MS, Allen S (2009) Audiovisual speech in older and younger adults: integrating a distorted visual signal with speech in noise. *Exp Aging Res* 35:202–219.
- Grant KW, Seitz PF (2000) The use of visible speech cues for improving auditory detection of spoken sentences. *J Acoust Soc Am* 108:1197–1208.
- Gruber O, Müller T, Falkai P (2007) Dynamic interactions between neural systems underlying different components of verbal working memory. *J Neural Transm* 114:1047–1050.
- Hampson M, Peterson BS, Skudlarski P, Gatenby JC, Gore JC (2002) Detection of functional connectivity using temporal correlations in MR images. *Hum Brain Mapp* 15:247–262.
- Hauser MD (1996) *The evolution of communication*. Cambridge, MA: MIT Press.
- Hein G, Knight RT (2008) Superior temporal sulcus— it's my area: or is it? *J Cogn Neurosci* 20:2125–2136.
- Helbig HB, Ernst MO (2008) Visual-haptic cue weighting is independent of modality-specific attention. *J Vis* 8:21 21–16.
- Hesselmann G, Sadaghiani S, Friston KJ, Kleinschmidt A (2010) Predictive coding or evidence accumulation? False inference and neuronal fluctuations. *PLoS ONE* 5:e9926.
- Hickok G, Poeppel D (2007) The cortical organization of speech perception. *Nat Rev Neurosci* 8:393–402.
- Horwitz B, Braun AR (2004) Brain network interactions in auditory, visual and linguistic processing. *Brain Lang* 89:377–384.
- Horwitz B, Warner B, Fitzer J, Tagamets MA, Husain FT, Long TW (2005) Investigating the neural basis for functional and effective connectivity. Application to fMRI. *Philos Trans R Soc Lond B Biol Sci* 360:1093–1108.
- Ihaka R, Gentleman R (1996) R: a language for data analysis and graphics. *J Comput Graph Stat* 5:299–314.
- Kanwisher N, Yovel G (2006) The fusiform face area: a cortical region specialized for the perception of faces. *Philos Trans R Soc Lond B Biol Sci* 361:2109–2128.
- Kersten D, Mamassian P, Yuille A (2004) Object perception as Bayesian inference. *Annu Rev Psychol* 55:271–304.
- Kreifelts B, Ethofer T, Grodd W, Erb M, Wildgruber D (2007) Audiovisual integration of emotional signals in voice and face: an event-related fMRI study. *Neuroimage* 37:1445–1456.
- Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci* 12:535–540.
- Kryter KD (1996) *Handbook of hearing and the effects of noise*. New York: New York Academic.
- Lee J, Maunsell JH (2009) A normalization model of attentional modulation of single unit responses. *PLoS ONE* 4:e4651.
- 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.
- Ma WJ, Zhou X, Ross LA, Foxe JJ, Parra LC (2009) Lip-reading aids word recognition most in moderate noise: a Bayesian explanation using high-dimensional feature space. *PLoS ONE* 4:e4638.
- Macaluso E, Frith C, Driver J (2000) Selective spatial attention in vision and touch: unimodal and multimodal mechanisms revealed by PET. *J Neurophysiol* 83:3062–3075.
- AQ: M
- AQ: N

- MacDonald J, Andersen S, Bachmann T (2000) Hearing by eye: how much spatial degradation can be tolerated? *Perception* 29:1155–1168.
- AQ: O** Mazziotta J, Toga A, Evans A, Fox P, Lancaster J, Zilles K, Woods R, Paus T, Simpson G, Pike B, Holmes C, Collins L, Thompson P, MacDonald D, Jacoboni M, Schormann T, Amunts K, Palomero-Gallagher N, Geyer S, Parsons L, et al. (2001) A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philos Trans R Soc Lond B Biol Sci* 356:1293–1322.
- McIntosh AR, Gonzalez-Lima F (1994) Structural equation modeling and its application to network analysis in functional brain imaging. *Hum Brain Mapp* 2:2–22.
- Miller LM, D'Esposito M (2005) Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *J Neurosci* 25:5884–5893.
- Mozolic JL, Joyner D, Hugenschmidt CE, Peiffer AM, Kraft RA, Maldjian JA, Laurienti PJ (2008) Cross-modal deactivations during modality-specific selective attention. *BMC Neurol* 8:35.
- Munhall KG, Kroos C, Jozan G, Vatikiotis-Bateson E (2004) Spatial frequency requirements for audiovisual speech perception. *Percept Psychophys* 66:574–583.
- Murray SO, Kersten D, Olshausen BA, Schrater P, Woods DL (2002) Shape perception reduces activity in human primary visual cortex. *Proc Natl Acad Sci U S A* 99:15164–15169.
- Narain C, Scott SK, Wise RJS, Rosen S, Leff A, Iversen SD, Matthews PM (2003) Defining a left-lateralized response specific to intelligible speech using fMRI. *Cereb Cortex* 13:1362–1368.
- Noesselt T, Rieger JW, Schoenfeld MA, Kanowski M, Hinrichs H, Heinze HJ, Driver J (2007) Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *J Neurosci* 27:11431–11441.
- AQ: P** Noppeney U, Josephs O, Hocking J, Price CJ, Friston KJ (2008) The effect of prior visual information on recognition of speech and sounds. *Cereb Cortex* 18:598–609.
- Obleser J, Wise RJS, Dresner MA, Scott SK (2007) Functional integration across brain regions improves speech perception under adverse listening conditions. *J Neurosci* 27:2283–2289.
- Olman CA, Ugurbil K, Schrater P, Kersten D (2004) BOLD fMRI and psychophysical measurements of contrast response to broadband images. *Vision Res* 44:669–683.
- Park JC, Zhang X, Ferrera J, Hirsch J, Hood DC (2008) Comparison of contrast-response functions from multifocal visual-evoked potentials (mfVEPs) and functional MRI responses. *J Vis* 8:1–12.
- Patel RS, Bowman FD, Rilling JK (2006) Determining hierarchical functional networks from auditory stimuli fMRI. *Hum Brain Mapp* 27:462–470.
- Patterson RD, Johnsrude IS (2008) Functional imaging of the auditory processing applied to speech sounds. *Philos Trans R Soc Lond B Biol Sci* 363:1023–1035.
- AQ: Q** Pelphrey KA, Morris JP, Michelich CR, Allison T, McCarthy G (2005) Functional anatomy of biological motion perception in posterior temporal cortex: an fMRI study of eye, mouth and hand movements. *Cereb Cortex* 15:1866–1876.
- Poeppel D, Idsardi WJ, van Wassenhove V (2008) Speech perception at the interface of neurobiology and linguistics. *Philos Trans R Soc Lond B Biol Sci* 363:1071–1086.
- Price CJ (2000) The anatomy of language: contributions from functional neuroimaging. *J Anat* 197:335–359.
- Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL (2001) A default mode of brain function. *Proc Natl Acad Sci U S A* 98:676–682.
- Reynolds JH, Heeger DJ (2009) The normalization model of attention. *Neuron* 61:168–185.
- Saad ZS, Glen DR, Chen G, Beauchamp MS, Desai R, Cox RW (2009) A new method for improving functional-to-structural MRI alignment using local Pearson correlation. *Neuroimage* 44:839–848.
- Scott SK, Johnsrude IS (2003) The neuroanatomical and functional organization of speech perception. *Trends Neurosci* 26:100–107.
- Scott SK, Blank CC, Rosen S, Wise RJ (2000) Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123:2400–2406.
- Sekiyama K, Kanno I, Miura S, Sugita Y (2003) Auditory-visual speech perception examined by fMRI and PET. *Neurosci Res* 47:277–287.
- Seltzer B, Pandya DN (1994) Parietal, temporal, and occipital projections to cortex of the superior temporal sulcus in the rhesus monkey: a retrograde tracer study. *J Comp Neurol* 343:445–463.
- Seltzer B, Cola MG, Gutierrez C, Masee M, Weldon C, Cusick CG (1996) Overlapping and nonoverlapping cortical projections to cortex of the superior temporal sulcus in the rhesus monkey: double anterograde tracer studies. *J Comp Neurol* 370:173–190.
- Shahin AJ, Miller LM (2009) Multisensory integration enhances phonemic restoration. *J Acoust Soc Am* 125:1744–1750.
- Shannon RV, Zeng F-G, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. *Science* 270:303–304.
- Smiley JF, Hackett TA, Ulbert I, Karmas G, Lakatos P, Javitt DC, Schroeder CE (2007) Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. *J Comp Neurol* 502:894–923.
- Stein BE, Meredith MA (1993) *The merging of the senses*. Cambridge, MA: MIT Press.
- Stein JL, Wiedholz LM, Bassett DS, Weinberger DR, Zink CF, Mattay VS, Meyer-Lindenberg A (2007) A validated network of effective amygdala connectivity. *Neuroimage* 36:736–745.
- Stevenson RA, James TW (2009) Audiovisual integration in human superior temporal sulcus: inverse effectiveness and the neural processing of speech and object recognition. *Neuroimage* 44:1210–1223.
- Stevenson RA, Kim S, James TW (2009) An additive-factors design to disambiguate neuronal and areal convergence: measuring multisensory interactions between audio, visual and haptic sensory streams using fMRI. *Exp Brain Res* 198:183–194.
- Sumbly WH, Pollack I (1954) Visual contribution to speech intelligibility in noise. *J Acoust Soc Am* 26:212–215.
- Thomas SM, Jordan TR (2002) Determining the influence of Gaussian blurring on inversion effects with talking faces. *Percept Psychophys* 64:932–944. **AQ: R**
- Tootell RB, Reppas JB, Kwong KK, Malach R, Born RT, Brady TJ, Rosen BR, Belliveau JW (1995) Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *J Neurosci* 15:3215–3230.
- Upadhyay J, Silver A, Knaus TA, Lindgren KA, Ducros M, Kim D-S, Tager-Flusberg H (2008) Effective and structural connectivity in the human auditory cortex. *J Neurosci* 28:3341–3349. **AQ: S**
- van Atteveldt NM, Blau VC, Blomert L, Goebel R (2010) fMR-adaptation indicates selectivity to audiovisual content congruency in distributed clusters of human superior temporal cortex. *BMC Neurosci* 11:11.
- van Atteveldt N, Roebroek A, Goebel R (2009) Interaction of speech and script in human auditory cortex: insights from neuro-imaging and effective connectivity. *Hear Res* 258:152–164.
- Van Essen DC (2005) A population-average, landmark- and surface-based (PALS) atlas of human cerebral cortex. *Neuroimage* 28:635–662.
- von Kriegstein K, Dogan O, Grüter M, Giraud AL, Kell CA, Grüter T, Kleinschmidt A, Kiebel SJ (2008) Simulation of talking faces in the human brain improves auditory speech recognition. *Proc Natl Acad Sci U S A* 105:6747–6752.
- Werner S, Noppeney U (2010) Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. *Cereb Cortex* 20:1829–1842.
- Wilson M (1988) *The MRC psycholinguistic database: machine readable dictionary, version 2*. *Behav Res Methods Instrum Comput* 20:6–11.
- Winer JA (2006) Decoding the auditory corticofugal systems. *Hear Res* 212:1–8.
- Wolmetz M, Poeppel D, Rapp B (2011) What does the right hemisphere know about phoneme categories? *J Cogn Neurosci* 23:552–569. **AQ: T**
- Wright TM, Pelphrey KA, Allison T, McKeown MJ, McCarthy G (2003) Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cereb Cortex* 13:1034–1043.
- Zatorre RJ (2007) There's more to auditory cortex than meets the ear. *Hear Res* 229:24–30.

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

1

- A—AU: Journal style is to define all nonstandard abbreviations at first use in title, abstract, and main text. The abbreviation is deleted if it is only used once in any of the above. Please verify throughout.
- B—AU: Per journal style, tables do not have parts A and B. I have separated Table 2 into Tables 2 and 3. Please verify edits, including citations. Please also verify added explanation of L in Table 3 footnote.
- C—AU: Per style, ‘that’ (with no comma) is used with restrictive clauses and ‘, which’ is used with nonrestrictive clauses. More generally, a phrase that is restrictive, that is, essential to the meaning of the noun it belongs to, should not be set off by commas. A nonrestrictive phrase should be enclosed in commas or, if at the end of a sentence, preceded by a comma. Please confirm correct usage throughout.
- D—AU: Please verify expansion of 2-AFC at only use.
- E—AU: Per journal style, italics are not allowed for emphasis.
- F—AU: Please verify change from ‘with either passive word presentation’ to ‘with passive word presentation ’ or correct as necessary.
- G—AU: Please verify edits to the sentence beginning “We define ‘noisy’ and ‘unreliable’ stimuli”. Originally, the definition provided for adjectives was a noun, so I added ‘stimuli’ and adjusted the definition to match.
- H—AU: Please check carefully and ensure that the code is set correctly throughout.
- I—AU: Most of the backslashes have a space in front of them. Four do not (two here and two below, preceded by ARevent and VRevent). Please correct if necessary.
- J—AU: Please verify added page number for Beauchamp et al. (2010a).
- K—AU: Please update Bernstein et al., 2010 with volume and page numbers, if possible. If not, please provide the date you retrieved the article and verify added doi.
- L—AU: Please verify edits to the Chen et al. (2007) reference.
- M—AU: Please verify added volume and page numbers for Ghose (2009).

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

2

N—AU: Please verify added 'e4638' for Ma et al. (2009).

O—AU: Per journal style, 20 authors are listed in references before using et al. Please verify added authors.

P—AU: Please verify changed year for Noppeney et al. (2008).

Q—AU: Please verify added volume and page numbers for Pelphrey et al. (2005).

R—AU: Please verify changed page numbers for Thomas and Jordan (2002).

S—AU: Please verify added second author for Upadhyay et al. (2008).

T—AU: Please verify added volume and page numbers and corrected year for Wolmetz et al. (2011).

U—AU: Please review all figure legends carefully and ensure that all abbreviations not defined in text are defined in the legends, and that all symbols, asterisks, and error bars are described accurately. Abbreviations defined in main text should not be redefined in legends.

V—AU: Please verify expansion of NSF and NIH at only uses in acknowledgments.
