5-2011

The Neural Substrates Of Multisensory Speech Perception

Audrey R. Nath

THE NEURAL SUBSTRATES OF MULTISENSORY SPEECH PERCEPTION

by

Audrey Rosa Nath, B.S., B.A.

APPROVED:

_____
Michael Beauchamp, Ph.D.
Supervisory Professor

_____
Ruth Heidelberger, M.D., Ph.D.

_____
Anne Sereno, Ph.D.

_____
Sandeep Agarwal, M.D., Ph.D.

_____
Tatiana Schnur, Ph.D.

APPROVED:

_____
Dean, The University of Texas
Graduate School of Biomedical Sciences

THE NEURAL SUBSTRATES OF MULTISENSORY SPEECH PERCEPTION

A DISSERTATION

Presented to the Faculty of The University of Texas Health Science Center at Houston

and The University of Texas M. D. Anderson Cancer Center

Graduate School of Biomedical Sciences

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

by

Audrey Rosa Nath, B.S., B.A., Houston, Texas

May, 2011

DEDICATION

This thesis is dedicated to my brother, Eugene Ravindra Nath (1978-2004), to whom I owe my childhood, my curiosity, and every last one of our jokes.

ACKNOWLEDGMENTS

I would like to thank my supervisory professor, Michael Beauchamp, for guiding me through every experiment, analysis, success and failure that I encountered in grad school. To my labmates, thank you for creating a wonderfully friendly lab environment that I will miss coming in to everyday.

I thank the faculty of my advisory, examining and supervisory committees for providing astute advice and pointing out flaws in my experiments and analyses long before I heard about them from peer reviewers. Thanks to Ruth Heidelberger, Sandeep Agarwal, Tatiana Schnur, Anne Sereno, Anthony Wright, Edward Jackson, Tim Ellmore and Jack Waymire for coming to all of those meetings.

I would like to thank my funding sources, the Center for Clinical and Translational Sciences (CCTS) T32 and the "Training in Neurosciences" T32 fellowships, for financial support as well as for guidance with professional development.

The MD/PhD Program here at UT-Houston has been just as much a strong community for me as well as a great source of support from faculty and students alike. I would like to thank my mentor within the MD/PhD Program, Ruth Heidelberger, for all of our talks about medicine, research and life. Thanks to Amy Reid for running all those marathons with me, to Rene Colorado, Chirag Patel and Katrina Salazar for enduring friendships, and to Shohrae Hajibashi (1979-2011) for inspiring us all.

Finally, I would not be the person I am today without my family. Mom and Dad, thank you for your unwavering support growing up and throughout this never-ending schooling—I owe any successes to both of you believing in me. And to my husband, Mark Flaum, thank you for listening.

THE NEURAL SUBSTRATES OF MULTISENSORY SPEECH PERCEPTION

Publication No. 1146

Audrey Rosa Nath, B.S., B.A.

Supervisory Professor: Michael Beauchamp, Ph.D.

Comprehending speech is one of the most important human behaviors, but we are only beginning to understand how the brain accomplishes this difficult task. One key to speech perception seems to be that the brain integrates the independent sources of information available in the auditory and visual modalities in a process known as multisensory integration. This allows speech perception to be accurate, even in environments in which one modality or the other is ambiguous in the context of noise. Previous electrophysiological and functional magnetic resonance imaging (fMRI) experiments have implicated the posterior superior temporal sulcus (STS) in auditory-visual integration of both speech and non-speech stimuli. While evidence from prior imaging studies have found increases in STS activity for audiovisual speech compared with unisensory auditory or visual speech, these studies do not provide a clear mechanism as to how the STS communicates with early sensory areas to integrate the two streams of information into a coherent audiovisual percept. Furthermore, it is currently unknown if the activity within the STS is directly correlated with strength of audiovisual perception. In order to better understand the cortical mechanisms that underlie audiovisual speech perception, we first studied the STS activity and connectivity during the perception of speech with auditory and visual components of

varying intelligibility. By studying fMRI activity during these noisy audiovisual speech stimuli, we found that STS connectivity with auditory and visual cortical areas mirrored perception; when the information from one modality is unreliable and noisy, the STS interacts less with the cortex processing that modality and more with the cortex processing the reliable information. We next characterized the role of STS activity during a striking audiovisual speech illusion, the McGurk effect, to determine if activity within the STS predicts how strongly a person integrates auditory and visual speech information. Subjects with greater susceptibility to the McGurk effect exhibited stronger fMRI activation of the STS during perception of McGurk syllables, implying a direct correlation between strength of audiovisual integration of speech and activity within an the multisensory STS.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Speech is a prevalent form of communication for humans, and understanding speech in noisy environments is a common task that people perform with relative ease. Under everyday conditions, we generally have access to both visual face movements and auditory vocal features that together aid in comprehension, making audiovisual speech perception a common occurrence of multisensory integration. Prior studies of audiovisual speech comprehension have shown that providing visual information helps subjects understand speech in the presence of noise (1-3). For example, when carrying out a conversation in a crowded restaurant, we focus on the speaker's mouth movements in order to decipher words in the midst of nearby conversations and background music.

Behavioral studies of multisensory integration in audiovisual speech have shown clear evidence of improved speech perception when the auditory and visual components of speech are presented together. A number of studies have found that speech perception is better for audiovisual speech than auditory speech alone. The presentation of visual mouth movements is known to improve comprehension of noisy auditory speech (1, 4-6). MacLeod and Summerfield (1990) found an 11-decibel "benefit" of visual speech in conjunction with low signal-to-noise auditory speech. Additionally, studies have found that speech perception is better for audiovisual speech than for visual speech alone. Risberg and Lubker (1978) found that subjects with normal hearing correctly perceived only 37.9% of test sentences when relying on visual speech-reading alone. When the subjects were presented with a low signal-to-noise version of the speech sound along with the corresponding visual mouth movements, performance jumped to 78.5% correctly perceived sentences. More recently, Remez, Fellowes, Pisoni, Goh and Rubin (1998) examined accuracy in identifying audiovisual sentences with clear video and

degraded sound. When subjects viewed the speaker's face without any sound, they identified sentences with an accuracy of 26.2%. Adding matching sine-wave sentences with the video, however, increased performance to 84.0%.

In different speech environments, auditory and visual noise levels can vary, resulting in changing reliabilities of the auditory and visual modalities. For instance, in a loud room, the auditory information is less reliable, while in a dark room, the visual information is less reliable. The integration of auditory and visual components of speech that has different levels of reliability has been found to follow the idea of optimal integration, in which the more reliable modality has greater influence on the behavioral decision (7-10). Alais and Burr (2004) tested the idea of optimal integration by presenting auditory clicks and visual blobs of varying widths to adjust visual reliability. The subjects were asked to identify the location of a simultaneous but spatially-misaligned pairing of the click and blob. The authors found that as the visual blob was smaller (and hence more reliable), localization of the audiovisual stimulus was dominated by the location of the visual stimulus. Conversely, when the visual blob was larger and less reliable, the localization was dominated by the location of the auditory stimulus. Ma et al. (2009) present a model of optimal integration in which auditory and visual inputs are represented as distributions in high-dimensional feature space. As the reliability of an input increases, the variance of its distribution decreases. The multisensory estimate of the word is then between the auditory and visual distributions but closer to the smaller distribution of the more reliable modality. In a study of subjects who were presented with incongruent audiovisual words of varying auditory reliability,

they found that low auditory reliability increased reports of the visual word while high auditory reliability increased reports of the auditory word.

While it has been shown that having both auditory and visual components of speech improves comprehension, and that audiovisual perception more closely follows the speech information presented in the more reliable (and less noisy) modality, one striking example of audiovisual integration shows a *changed* perception of speech when both modalities are present. McGurk and MacDonald (11) showed a remarkable example of audiovisual integration for clear spoken syllables; an auditory "ba" presented with the mouth movements of "ga" is perceived by the listener as "da." Here, multisensory integration is apparent given the perception of a third, distinct syllable separate from either syllable perceived in the auditory or visual modality. This finding by McGurk and MacDonald emphasizes that audiovisual integration is more than a small mechanism to aid speech perception in noise, but a powerful effect worthy of independent investigation.

Therefore, both the auditory and visual components of speech are very important in speech comprehension, and these two streams of information must be integrated together within the cerebral cortex. The auditory speech information is processed in Heschl's gyrus, planum temporale and associated auditory cortical regions in the superior temporal gyrus (12, 13) and is further processed in anterior and posterior portions of the superior temporal sulcus (STS), inferior frontal, temporo-parietal and inferior temporal structures (14-18). Visual speech information is processed in the visual pathway starting from early visual areas in occipital cortex (19-21) and is further processed in posterior STS as well as inferior frontal gyrus (IFG) and premotor cortex

(22-24). The auditory and visual information are hence processed jointly in higher-order areas in inferior frontal areas in and around Broca's area, a region important for speech production (25-28), and posterior STS within Wernicke's area, a region important for speech comprehension (29, 30).

As such, cortical areas within both the STS and IFG are important for language processing, speech perception and multisensory integration (14, 31-35). For example, inferior frontal areas around Broca's area show differential activity for different types of audiovisual speech, with greater responses to incongruent than congruent audiovisual speech (26). Eisner et al. (2010) found greater activity in the left IFG during noisy words in subjects who were better able to learn to recognize these noisy words after training. Other studies have found differential activity in temporal areas, including STS, which was correlated with individual language abilities. Wong et al. (2007) found that areas of the left posterior STS showed increased activation in subjects who more readily acquired tone patterns in a novel tone-based language, while right-sided areas including the right posterior STS showed increased activation in the subjects who had more difficulty in learning these pitch patterns. Similarly, Mei et al. (2008) found increased activity in left middle temporal gyrus and STS in Chinese speakers who better able to learn an artificial language.

While both the auditory and visual speech information is processed in inferior frontal areas as well as posterior STS, only the left posterior STS shows consistently greater activation during audiovisual speech stimuli as opposed to auditory or visual speech alone (20, 36-44). This multisensory region of the STS is anatomically connected to both auditory and visual areas in the cortex (45, 46). Electrophysiological studies of

both macaques (36, 37) and humans (38) have identified the posterior STS as a site of multisensory integration of audiovisual speech. Ghazanfar, Chandrasekaran and Logothetis (2008) recorded local field potentials from left STS and auditory cortex of two rhesus monkeys during presentation of auditory, visual and audiovisual monkey vocalizations. The functional interactions between the auditory cortex and the STS increased in strength during presentations of dynamic faces and voices relative to either communication signal alone. Kayser and Logothetis (2009) studied effective connectivity between neurons in auditory cortex and STS in macaques during auditory, visual and audiovisual movies of animals and cartoons. They found that multisensory regions of auditory cortex received stronger feedback from STS during audiovisual stimuli than during auditory-only stimuli, while regions of auditory cortex exhibiting multisensory suppression received weaker feedback. Reale et al. (2007) measured event related potentials (ERPs) from subjects undergoing electrode implantation surgery as part of management for intractable epilepsy. These subjects were presented with auditory, visual, congruent audiovisual and incongruent audiovisual syllables as ERPs were recorded from posterior lateral superior temporal gyrus. They found that visual facial information either heightened or decreased the auditory signal from this area, with a larger area showing this effect in the language-dominant hemisphere.

Functional magnetic resonance imaging (fMRI) studies of multisensory integration in speech have found an increased activation of left STS in response to audiovisual speech as opposed to either modality presented alone (20, 39-44). Callan et al. (2004) examined fMRI activity during audiovisual speech consisting of multispeaker auditory noise and congruent visual sentences at three levels of visual noise. The

strongest sites of multisensory integration were in left middle temporal gyrus and left superior temporal gyrus and sulcus. Sekiyama, Kanno, Miura and Sugita (2003) examined the McGurk effect with clear video and low signal-to-noise auditory input. They found that as the auditory speech became noisier, subjects exhibited a stronger McGurk effect, i.e. they had fewer responses corresponding to the auditory syllable. Similarly, they found greater activation in the STS for the audiovisual McGurk stimuli with lower auditory intelligibility than the stimuli with high intelligibility. Stevenson and James (2009) found that for both speech and tool congruent audiovisual stimuli, both auditory and visual components needed to be degraded in order to achieve a STS response which was greater for audiovisual stimuli than for the sum of the responses to its auditory and visual components, also known as a superadditive response. Calvert et al. (2000) found a superadditive response to congruent audiovisual speech in the STS that was not present during incongruent audiovisual speech. The result implied that congruency of audiovisual speech is sufficient for a multisensory, superadditive response in STS, though this finding was not replicated in Stevenson and James (2009). Miller and D'Esposito (2005) found that congruent audiovisual speech activates posterior STS more than incongruent (mismatching) speech, though not necessarily in a superadditive manner. Beauchamp et al. (2004) studied STS activation during presentation of auditory, visual and audiovisual tools and speech. They found that within the STS, there were smaller areas that had predominantly auditory, visual or audiovisual activity. Their findings suggest that information from different modalities is brought to the STS separately and then integrated.

In a recent study using transcranial magnetic stimulation (TMS) directed to the posterior STS, this method of virtual lesioning provided evidence that activity within the STS is necessary for the integration of auditory and visual components of speech (47). To examine if audiovisual integration is disrupted when neural firing in the STS is interrupted, the authors studied perception of the audiovisual McGurk effect during TMS directed to the STS, TMS directed to a control site, or no delivery of a TMS pulse. TMS directed to the STS was found to decrease the perception of the McGurk illusion from 94% of trials without TMS to 43% of trials with TMS. These results signify that audiovisual integration of speech depends upon the uninterrupted activity of neurons within the multisensory STS. Given the difficulty in finding patients who have selective lesions to the posterior STS and our inability to create permanent targeted lesions in humans, this result provides the first evidence from a targeted lesioning study that shows the necessary role of the STS in audiovisual integration of speech.

While evidence from prior studies have found increases in STS activity for audiovisual speech compared with unisensory auditory or visual speech, these studies do not provide a clear mechanism as to how the STS communicates with early auditory and visual cortical areas to integrate the two streams of information into a coherent audiovisual percept. Furthermore, it is currently unknown if the activity within the STS is directly correlated with strength of audiovisual perception. The goals of my project are two-fold:

1. To elucidate the mechanism for integration of auditory and visual speech by studying functional connectivity between STS and auditory and visual cortical areas (Chapter 2).

2. To characterize the response of the STS during audiovisual speech and determine if activity within the STS serves as a predictor for strength of perception of McGurk syllables (Chapter 3).

We first elucidated the mechanism by which the STS integrates information from connected auditory and visual areas. We predicted that strengths of connection from sensory areas to multisensory STS should mirror perception of audiovisual stimuli under the rules of optimal integration: when information from one modality is unreliable and noisy, the STS should interact less with the cortex processing that modality and more with the cortex processing the reliable information. For example, if audiovisual speech is unreliable in the auditory modality and reliable in the visual modality (i.e. noisy auditory component and clear visual mouth movements), then perception should more closely resemble what was presented in the more reliable visual modality, and functional connections from visual cortex to STS should be stronger than those from auditory cortex to STS. Conversely, if audiovisual speech is reliable in the auditory modality and unreliable in the visual modality (i.e. clear auditory component and blurry visual mouth movements), then perception should more closely resemble what was presented in the more reliable auditory modality, and functional connections from auditory cortex to STS should be stronger than those from visual cortex to STS.

We then characterized the role of STS activity during varying audiovisual speech perception to determine if activity within the STS predicts how strongly a person integrates auditory and visual speech information. In order to clarify how brain activity within an individual's STS correlates with that person's audiovisual perception, we studied the amplitude of cortical response within the STS as measured by fMRI during

the audiovisual McGurk illusion as well as during syllables not associated with any audiovisual illusion. We hypothesized that subjects who perceive the McGurk illusion more strongly will have a correlated increase in amplitude of response of the multisensory STS.

CHAPTER 2: CORTICAL CONNECTIVITY DURING NOISY AV SPEECH

Introduction

Humans understand speech by combining the independent sources of information available in the auditory and visual modalities, making speech perception an important example of multisensory integration (2, 3, 11). The perceptual and neural benefits of multisensory integration are most pronounced when input stimuli are weak (48), a property that can be quantified as reliability, the variability in the physical and neural representation of the stimulus (49). The reliability of speech information differs across environments: in a loud room, auditory information is less reliable, while in a dark room, visual information is less reliable. Behavioral experiments have shown that for both speech and non-speech stimuli, subjects are biased towards perceiving the stimulus presented in the reliable modality, a phenomenon termed reliability-weighting (7-10).

Although behavioral reliability weighting is a widespread mechanism for dealing with dynamically changing noise in the input modalities to multisensory integration, little is known about how the brain performs this process. In one model of reliability-weighted multisensory integration, the Bayesian cue integration model, the brain weights information from the early sensory input areas into the multisensory brain areas depending on how reliable that modality is. As described recently by Ma et al. (2009), auditory and visual speech inputs are represented as distributions in high-dimensional feature space. As the reliability of an input increases, the variance of its distribution decreases. The multisensory estimate of the word is then between the auditory and visual distributions but closer to the smaller distribution of the more reliable modality.

A brain area likely to mediate this multisensory function for audiovisual speech is a region in human posterior superior temporal sulcus (STS) which is known for

integrating auditory and visual information about speech and non-speech stimuli (20, 39-44, 50). In macaque STS, a region known as STP (superior temporal polysensory) or TPO (temporo-parietal-occipital) receives projections from auditory and visual association cortex (45, 46) and contains single neurons that show enhanced responses to auditory and visual communication signals (51). For brevity, we refer to the human homolog of this region as "STS" while noting that the STS also contains other functionally and anatomically heterogeneous regions (52-54). 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 (14, 31, 32, 55-59).

While it is clear that pSTS is involved with multisensory decision making through connections to early sensory areas, it is important to distinguish anatomical connections from functional connectivity between brain regions. The anatomical connections between visual and auditory cortex and STS exist at all times, but their strength and directionality can change under difference circumstances. Functional connectivity measures how closely two neuronal activities match each other over time during a particular cognitive state, inferring strength of interaction between two regions (60).

Our hypothesis is that the strengths of connections between STS and early sensory areas underlie the behavioral phenomenon of reliability weighting and should be modulated based on the reliability of the stimulus in that modality. We first created audiovisual speech stimuli of varying reliability: auditory-reliable stimuli consisted of clear auditory input with blurred visual input, while visual-reliable stimuli consisted of

blurred auditory input with clear visual input. We then established behavioral reliability weighting with auditory-reliable and visual-reliable speech stimuli. The amplitude of the neural response was measured using fMRI, and functional connectivity between sensory areas and STS was measured with structural equation modeling (60-64) in order to determine whether changes in amplitude or connection weights accompany reliability-weighted processing of speech.

Methods

Subjects and Stimuli

Thirty-four healthy subjects (thirteen female, mean age 27.6; ten subjects in Experiment 1, ten in Experiment 2, six in Experiment 3, six in Experiment 4, ten in Experiment 5) 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. All subjects were right-handed and did not have any visual or hearing impairments.

The auditory, visual and audiovisual stimuli created for each experiment are summarized in Table 2.1. Four stimulus classes were presented in separate experiments: auditory words, silent word videos, audiovisual words with degraded visual component (auditory-reliable) and audiovisual words with degraded auditory component (visual-reliable). The word stimuli consisted of 160 words from the MRC Psycholinguistic Database with imageability rating greater than 100, Brown verbal frequency of 20 to 200, age of acquisition less than seven years and Kucera-Francis written frequency greater than 80 (65). Each word and syllable was spoken by a female speaker, and the resulting audiovisual recordings were about 1 second long. White poster board was used as a backdrop and ceiling lamps provided lighting. Lighting was positioned to minimize asymmetric shadowing on the face and ambient noise was minimized. The duration of the words ranged from 1.1 to 1.8 seconds with ISI occupying the remainder of each 2-second trial.

	fMRI Expt Design	**Auditory-only**	**Visual-only**	**Auditory-Reliable**	**Visual-Reliable**	**Task**
Functional Localizer	Blocked	Undegraded Words (C)	Undegraded Words (C)	n/a	n/a	passive
Expt 1	Blocked	*n/a*	*n/a*	Words (C)	Words (C)	passive
Expt 2	Event-Related	*n/a*	*n/a*	Words (C)	Words (C)	passive
Expt 3	Event-Related	*n/a*	*n/a*	n/a	Undegraded (C+I) Mid-blur (C+I) High-blur (C+I)	C vs. I C vs. I C vs. I
Expt 4	Event-Related	*n/a*	*n/a*	Syllables (C)	Syllables (C)	Attn-A: "Ja" vs. "Ma" Attn-V: Eyes open vs. closed
Expt 5	n/a	*n/a*	*n/a*	Syllables (C+I)	Syllables (C+I)	"Ma" vs. "Na"

Table 2.1 Stimuli and tasks

Expt: Experiment, C: congruent; I: incongruent; Attn-A: auditory attention; Attn-V: visual attention; n/a: not applicable for that experiment.

A schematic of the four stimulus types are shown in Figure 2.1. Auditory words were extracted from the recorded videos using QuickTime converter to .wav file format, 48 kHz rate, sample size 16 bits and normal render quality. The auditory-only stimuli consisted of the auditory portion of the speech and white crosshairs in the visual modality. Visual, silent movie clips of words were extracted from the recorded videos using QuickTime converter to .avi file format using the DV/DVCPRO –NTSC codec, 4:3 aspect ratio and interlaced scan mode. Visual-only words consisted of silence in the auditory modality and the visual mouth movements in the visual modality, followed by a scrambled image for 50 milliseconds in order to minimize afterimages. The baseline condition consisted of silence in the auditory modality and fixation crosshairs in the visual modality.

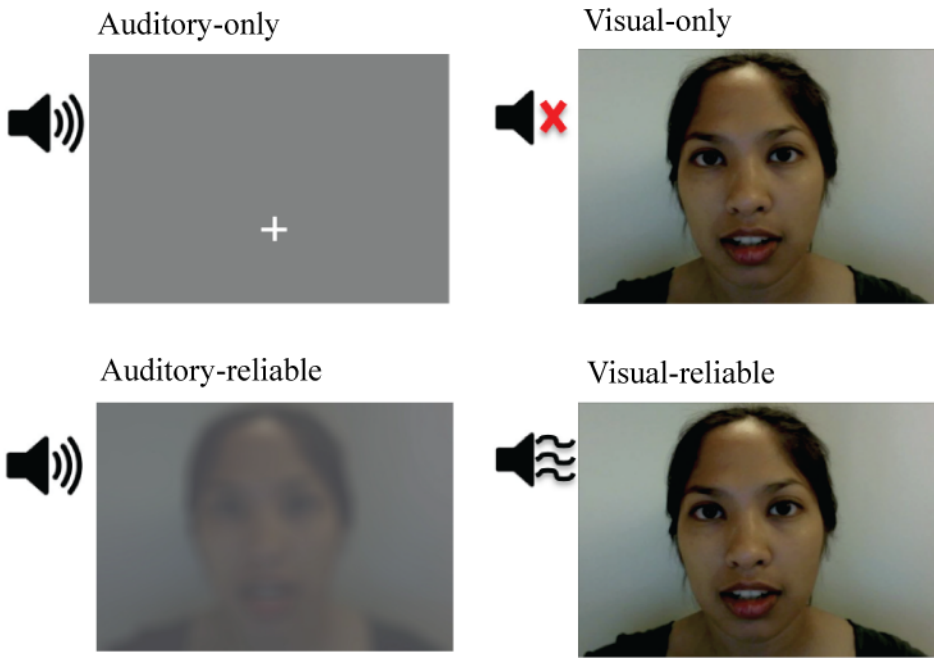

Figure 2.1 Auditory and visual stimuli

A. Undegraded auditory speech (loudspeaker icon) with visual fixation crosshairs.
B. Undegraded visual speech (illustrated by a single frame from a video) with no auditory stimulus.
C. Undegraded auditory with degraded video: auditory-reliable speech.
D. Undegraded video with degraded auditory: visual-reliable speech.

The reliability of the multisensory words was manipulated by modifying the auditory and visual components of the speech recordings. The auditory speech was degraded using Matlab (Mathworks, Inc.) with a noise-vocoded filter (66). The resulting noise-vocoded speech consisted of noise within the same temporal envelope of the original stimulus. As in Shannon et al. (1995), four separate temporal envelopes containing noise were created in four frequency bands: 1) 0-800 Hz, 2) 800-1500 Hz, 3) 1500-2500 Hz and 4) 2500-4000 Hz. The waveforms were downsampled at a smoothing frequency of 300 Hz. This method of noise-vocoding has been found to decrease intelligibility of auditory words (66). The visual component was degraded by first decreasing the contrast by 70% and then blurring the digital video with a Gaussian filter using Matlab. This method of decreasing the spatial resolution of visual speech stimuli has been found to decrease word identification (67).

Experiments 4 and 5 were performed using single syllables in which the auditory and visual reliability were manipulated. In Experiment 4, BOLD fMRI data was collected while subjects attended to either the visual or auditory modality. In Experiment 5, subjects made behavioral judgments about which of two syllables was perceived.

General fMRI Methods

Anatomical scans for each subject consisted of two T1-weighted scans anatomical collected at 3T using an 8-channel head gradient coil. The two anatomical scans were aligned, averaged into one dataset, transformed to the Talairach coordinate system (68). Each anatomical dataset was normalized to the the N27 reference anatomical volume (69) for group analysis. A three-dimensional cortical surface model

was created from these T1-weighted scans using FreeSurfer (70, 71), and functional data was overlaid onto this surface model using SUMA (72).

Functional scans consisted of T2*-weighted images collected using gradient-echo echo-planar imaging (TR = 2015 ms, TE = 30 ms, flip angle = 90°) with in-plane resolution of 2.75 x 2.75 mm. Thirty-three axial slices were collected at 3 mm intervals in order to collect data from the entire cerebral cortex. Each functional scan series consisted of 153 brain volumes. The first three volumes of each scan were discarded, resulting in 150 usable volumes.

Stimuli were presented using Presentation version 12 (Neurobehavioral Systems, Albany, CA). Auditory stimuli were presented to subjects within the scanner using MRI-compatible pneumatic headphones. Visual stimuli were projected onto a screen and subsequently viewed through a mirror attached to the head coil. Button presses were used to assess subject performance of tasks and were collected using a fiber-optic button response pad (Current Designs, Haverford, PA). An eye tracking system to ensure alertness and visual fixation during all functional scans (Applied Science Laboratories, Bedford, MA).

Analysis of the fMRI data was performed using Analysis of Functional NeuroImages software (AFNI) (73). The false discovery rate (FDR) procedure (74) was used to correct for multiple comparisons, and the FDR's were reported as "q" values. Functional activation was analyzed first within each individual subject, and then data was combined across subjects using a random-effects model. Functional activation maps were aligned to each subject's averaged anatomical scan and were 3-dimensionally motion-corrected using a local Pearson correlation (75).

A deconvolution analysis was performed for each subject to create functional activation maps using the AFNI function *3dDeconvolve* using a generalized linear model (76). In the experiments using a block design, one regressor was created for each stimulus type, and the time series of activation for each voxel in each scan was convolved with the stimulus timing of a boxcar-shaped, gamma-variate estimate of the hemodynamic response function for each regressor. In the experiments using a rapid event-related designs, one regressor was created for each individual presentation of each stimulus, and then a convolution was performed using –stim_times_IM mode of *3dDeconvolve* to estimate the amplitude of response to each individual stimulus. To help correct for head motion, six movement regressors were created for each scan and were modeled as regressors of no interest.

fMRI Functional Localizer and Regions of Interest

A functional localizer consisting of blocks of auditory and visual words was used to identify three regions of interest (ROIs) in each subject important for speech processing: auditory cortex, visual cortex, and STS (see Figure 2.2 for ROIs from individuals subjects and corresponding BOLD time series from each area). The ROIs were obtained from separate scan series, apart from the scan series for collecting audiovisual data, in order to prevent bias and avoid the phenomenon of "double-dipping" (77). Six ROIs were created for each subject, with three in the left hemisphere and three in the right hemisphere. Our main set of analyses used the ROIs created in the left hemisphere since the language-related activity is generally observed more in the left hemisphere (78, 79).

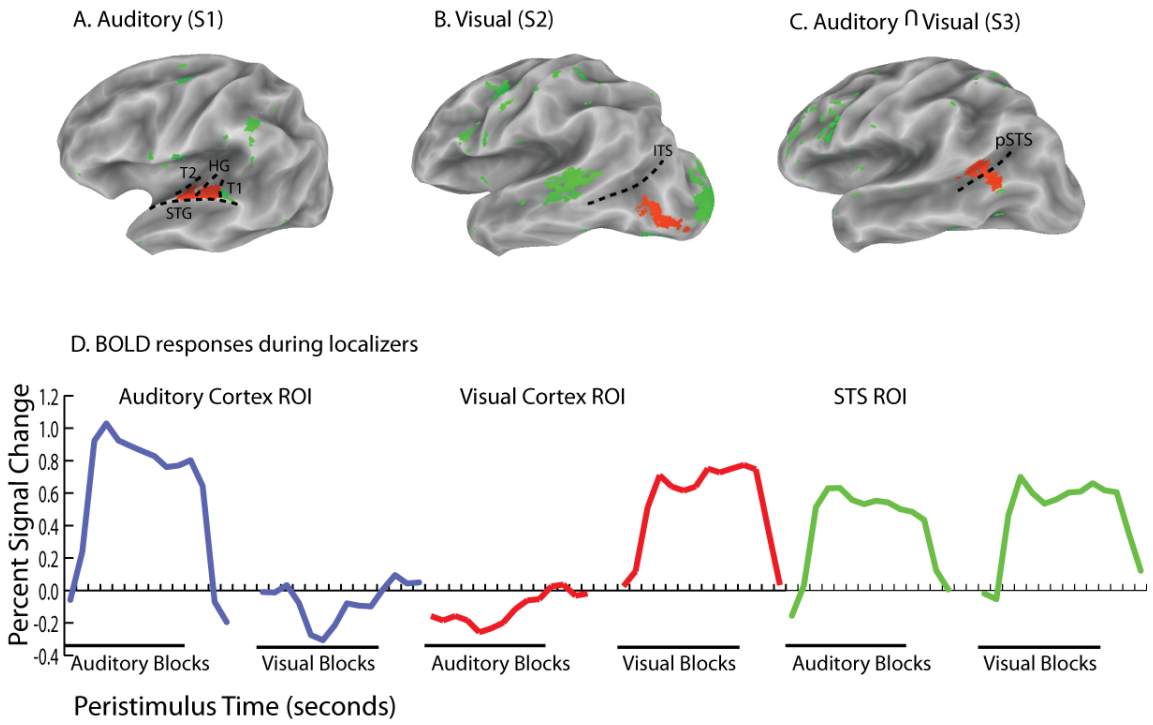

Figure 2.2 Regions of interest

A. Significant activation during the auditory fMRI localizer in subject S1 (orange: activity within the auditory ROI; green: activity outside the ROI). The dashed lines show the anatomical landmarks used to define the ROI. STG: crown of the superior temporal gyrus; T2: fundus of the second temporal sulcus; HG: crown Heschl's gyrus; T1: fundus of first temporal sulcus. Superior-lateral view of partially inflated left hemisphere.
B. Significant activation during visual fMRI localizer in subject S2 (orange: activity within the visual ROI; green: activity outside the ROI). ITS: posterior continuation of the inferior temporal sulcus. Lateral view of the partially-inflated left hemisphere.
C. Conjunction map of activation during auditory and visual fMRI localizers in subject S3 (orange: activity within the STS ROI; green: activity outside the ROI). pSTS: fundus of the posterior STS.
D. Timecourse of BOLD response during fMRI localizers. Auditory ROI curves in blue, visual ROI curves in red, and STS ROI curves in green. Within each set of ROI curves: response to blocks of auditory stimuli on left; response to blocks of visual stimuli on right. Black bar below x-axis shows stimulus block onset and offset.

The functional localizer contained five unisensory auditory and five unisensory visual blocks presented in random order. Each block ten trials (2 seconds each), one undegraded word per trial, and there were 10 seconds of fixation between each block. The auditory, visual and STS ROIs were created separately for each subject on the cortical surface. Voxels within the auditory ROI were chosen to center on Heschl's gyrus within boundaries for the primary auditory cortex based on prior work (80, 81). These boundaries consisted of the superior temporal gyrus in the lateral direction, the medial termination of Heschl's gyrus in the medial direction, the first temporal sulcus in the anterior direction and the transverse temporal sulcus in the posterior direction. Within these boundaries, voxels with activation greater than baseline during auditory-only blocks were used for further analysis. Voxels within the visual ROI were chosen to center within extrastriate lateral occipital cortex, a brain region critical for processing moving and biological stimuli which includes the middle temporal visual area and the extrastriate body area (82-87). Voxels with along the inferior temporal sulcus (ITS) or its posterior continuation near areas LO and MT (88). Within these boundaries, voxels with activation greater than baseline during visual-only blocks were used for further analysis. Voxels within the STS ROI were chosen within the anatomically-defined posterior STS for each subject (89, 90). Voxels with activity greater than baseline during both auditory-only and visual-only blocks were used for further analysis ($q < 0.05$ for each modality).

Structural Equation Modeling

Connection weights between auditory cortex, visual cortex and the STS during audiovisual stimuli were calculated using structural equation modeling. For each subject, a structural equation model consisted of connections between the STS and auditory and

visual ROIs (see Figure 2.3 for ROIs and model in one subject). The model was tested using both unidirectional and bidirectional connections as well as in both the left and right hemispheres. Path coefficients from the models were calculated using the software package "R" (91) and compared across subjects using an ANOVA.

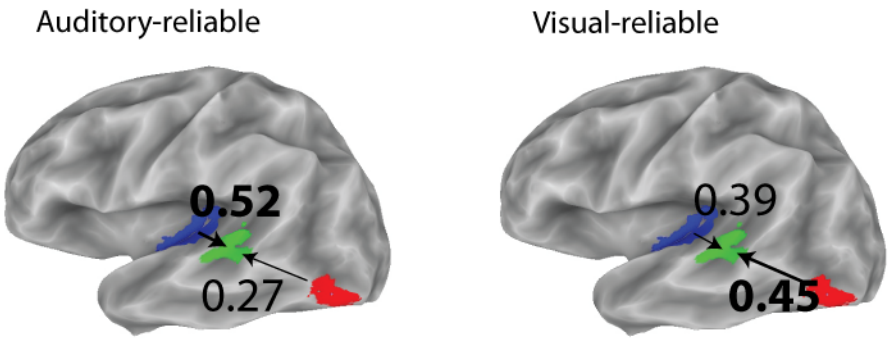

Figure 2.3 Functional connectivity in one subject

Functional connectivity between auditory cortex (blue), visual cortex (red) and STS (green) ROIs for one subject. Numbers indicate the path coefficients between the areas during perception of auditory-reliable speech (left) and visual-reliable speech (right). Lateral view of the partially inflated left hemisphere, dark gray shows sulcal depths, light gray shows gyral crowns.

Experiment 1: fMRI Block Design

Hemodynamic responses within each ROI and effective connectivity between the ROIs was investigated with a block design. To examine changes in connectivity for auditory-reliable and visual-reliable stimuli, one level of auditory unreliability and one level of visual unreliability were created. The auditory unreliable condition was created using a noise-vocoding procedure (see "Subjects and Stimuli") and the visual unreliability condition was created by blurring the videos with a 30-by-30 pixel Gaussian filter. Reliable auditory stimuli were paired with unreliable visual stimuli (auditory-reliable) and unreliable auditory stimuli were paired with reliable visual stimuli (visual-reliable). Each subject was presented with three 5-minute scan series of blocked stimuli, each containing five auditory-reliable and five visual-reliable blocks in random order with ten seconds of fixation baseline between each block. Stimulus blocks consisted of ten words each, with one different word per 2-second trial. The stimulus videos lasted between 1.1 to 1.8 seconds, and the interstimulus interval between word stimuli consisted of the fixation baseline.

As shown in Figure 2.2, blocks of auditory-reliable and visual-reliable stimuli evoked strong, boxcar-shaped hemodynamic responses. The overall shape of the square-shaped responses, however, would cause artificially high correlations between timeseries not because of similarities in activity for different stimuli, but because of the large change in amplitude for the onset and offset of each block. To remove this source of artifact, normalized time series were constructed by subtracting the amplitude of the mean response to each condition from the average time series, preventing the high-amplitude block onsets and offsets from artificially inflating the correlation between

ROIs (92). The path coefficients for the structural equation model were then calculated on these normalized time series, separately for each subject (93). For group analysis, within-subjects two-way ANOVAs were performed with stimulus reliability (auditory-reliable or visual-reliable) and sensory cortex (auditory or visual) as factors and amplitude of response and path coefficient as the dependent measures.

An additional whole brain psycho-physiological interaction (PPI) analysis was performed to search for other brain areas showing condition-dependent changes in connection strength with STS (94). The PPI analysis was performed with the STS time course as the physiological factor and stimulus type (auditory-reliable or visual-reliable) as the psychological factor. The PPI term was built by multiplying the STS time course with the psychological factor. The hemodynamic response of all voxels was deconvolved with the physiological factor, psychological factor and PPI terms as regressors. A random-effects group analysis was performed on the PPI contrasts ($T > 4$, $p < 0.01$). Spatial transformation to Talairach space was performed using the AfNI function *adwarp*. For each subject, the normalized STS time series was the physiological factor and stimulus condition (auditory-reliable or visual-reliable) was the psychological factor.

Experiment 2: fMRI Rapid Event-Related Design

Hemodynamic responses and effective connectivity were investigated with a rapid event-related (RER) design. Each subject was presented with two scan series, each containing sixty auditory-reliable words (2 s each), sixty visual-reliable words (2 s each) and thirty fixation baseline trials (2 s) presented pseudo-randomly in optimal rapid event-related order (95). The amplitude of response for each individual word stimulus

(sixty for each stimulus type) was obtained using deconvolution and averaged within each ROI. The input to the path analysis consisted of the response to each word in each ROI measured with deconvolution. The path coefficients were then entered into the group ANOVA.

Experiment 3: fMRI Rapid Event-Related Parametric Design

We next examined hemodynamic responses and effective connectivity using three levels of visual reliability in order to determine if step-wise changes in visual reliability resulted in a parametric changes in visual cortex BOLD amplitude and STS-visual cortex connectivity. The level of auditory reliability did not change from stimulus to stimulus; all auditory stimuli used the same parameters as the auditory-unreliable stimuli in experiments 1 and 2. There were four levels of increasing visual reliability examined: the most unreliable used a 30x30 Gaussian blur, intermediate levels using 5x5 and 15x15 Gaussian blurs, and the most reliable level using no blur at all (clear image). Each subject was presented with three scan series, each containing 30 presentations of each of the four stimulus types and 30 presentations of the baseline condition in pseudo-random order. Since there were no differences in the behavioral perception of the two intermediate blurring levels (5x5 and 15x15 Gaussian blurs) and no differences in the connectivity between these two levels, data from these two stimulus types were collapsed together for analysis.

Within each stimulus type, half of the trials were congruent and half were incongruent. Subjects had a 2-AFC task and responded with a button press if the perceived audiovisual word was congruent (same in auditory and visual modalities) or incongruent. As in experiment 2, one structural equation model consisting of three ROIs

with connections between auditory cortex and STS and visual cortex and STS was created and evaluated during the three levels of visual-reliability. For group analysis, within-subjects two-way ANOVAs were performed with visual stimulus reliability as a main factor and amplitude of response within the visual cortex and path coefficient between STS and visual cortex as the dependent measures.

Experiment 4: Attention Experiment

In order to determine if attention directed to one modality would enhance or override reliability-weighted connectivity changes, a rapid event-related design was used with congruent syllable stimuli ("ja" or "ma") that could be either auditory-reliable or visual-reliable. To direct attention to the auditory modality, subjects pressed a button to indicate the identity of each auditory syllable (if the syllable was "ja" or "ma"). To direct attention to the visual modality, subjects pressed a button to indicate the visual appearance of the speaker in the video (if the eyes were open or closed). We chose these tasks in order to maintain attention to each modality, and these tasks were kept relatively simple to avoid causing large task-related effects in the brain.

Each subject was presented with four scan series, two with the auditory attention task and two with the visual attention task. Eight stimulus types were constructed using a 2x2x2 design, with reliability (auditory-reliable or visual-reliable), syllable ("ja" or "ma") and appearance (eyes open or closed) as factors. Each scan series contained thirty presentations of each stimulus type (120 total) and thirty presentations of the baseline condition in a random order. The amplitude of response for each individual syllable stimulus (thirty for each stimulus type) was obtained using deconvolution and averaged within each ROI. The input to the path analysis consisted of the response to each word in

each ROI measured with deconvolution. The path coefficients were then entered into the group ANOVA.

Experiment 5: Behavioral Experiment

Subjects were presented with auditory-reliable and visual-reliable stimuli outside of the MR scanner to determine if these stimuli of varying reliability are perceived in a reliability-weighted manner. Eight stimulus types were constructed using a 2x2x2 factorial design: the first factor was auditory syllable ("ma" vs. "na"), the second factor was visual syllable ("ma" vs. "na") and the third factor was reliability (auditory-reliable vs. visual-reliable). Each of ten subjects was presented with 80 stimuli (10 examples of each stimulus type) and made a 2-AFC about each stimulus ("ma" vs. "na"). Responses to incongruent stimuli (e.g. auditory "na" paired with visual "ma") were analyzed with a within-subjects paired t-test.

Results

fMRI Localizer Experiment

The speech stimuli presented in the functional localizer scan series evoked robust hemodynamic responses in auditory cortex for auditory speech and in visual cortex for visual speech. The STS responded strongly to both auditory and visual speech (see Figure 2.2 for average BOLD timeseries from all ROIs and Table 2.2 for standard coordinates). The functional localizers were independent from the experimental scan series described below, allowing statistical tests to be performed without bias.

A. ROI locations

		Talairach Coordinates (mm)		
ROI	**Size (mm^3)**	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

B. Whole-brain connectivity analysis

			Talairach Coordinates		
Interaction	**Brain Region**	**Size (mm^3)**	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

Table 2.2 Locations of ROIs and activity in whole-brain analysis

A. Average size and location of individual auditory, visual and STS ROIs created from functional localizers across all subjects (mean +- SD).
B. Regions in the Experiment 1 group dataset showing a positive interaction with STS during auditory-reliable blocks or visual-reliable blocks.

Experiment 1: Block Design fMRI Experiment

Using a blocked design, we measured the brain response to two different types of speech: auditory-reliable words (auditory-reliable + visual-unreliable) and visual-reliable words (visual-reliable + auditory-unreliable). Two-way ANOVAs were performed with condition (auditory-reliable or visual-reliable) and sensory cortex (auditory or visual) as factors. Audiovisual words evoke a robust response from both auditory cortex and visual cortex; therefore, we did not expect a main effect of sensory cortex or of reliability. However, we predicted an interaction between reliability and sensory cortex, with the sensory cortex processing the reliable stimuli responding more strongly. As predicted, the ANOVA on the BOLD amplitude with ROI and stimulus condition as factors revealed a significant interaction ($F_{(1,9)} = 46.6$, $p = 0.00007$) driven by a greater BOLD response to auditory-reliable words in auditory cortex (Fisher's LSD test: $p_{LSD} < 0.0001$) and to visual-reliable words in visual cortex ($p_{LSD} < 0.05$; Figure 2.4; Table 2.3).

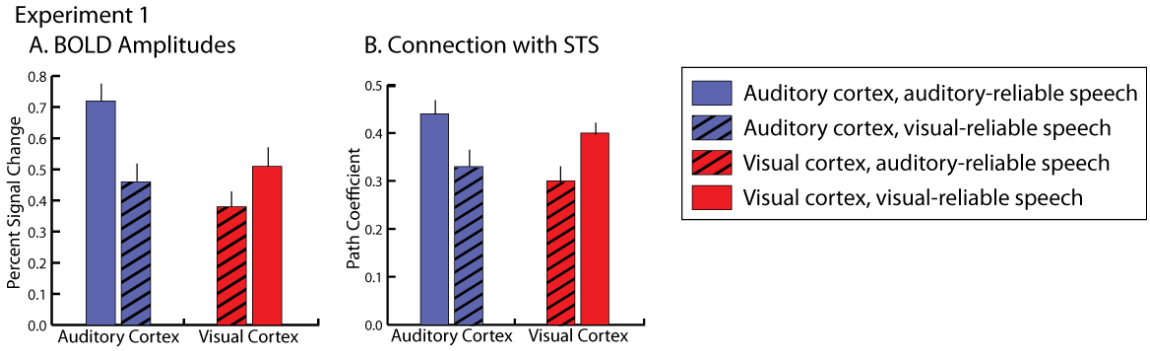

Figure 2.4 BOLD amplitudes and connection weights in Experiment 1

A. BOLD amplitudes in Experiment 1 reported as percent signal change. Error bars are standard error of the mean.
B. Connection weights between auditory and visual cortex and STS in Experiment 1 reported as path coefficients.

	Auditory Cortex	Visual Cortex	STS
Experiment 1			
Auditory-Reliable	0.72 +- 0.08	0.38 +- 0.07	0.68 +- 0.06
Visual-Reliable	0.46 +- 0.07	0.51 +- 0.07	0.61 +- 0.07
Experiment 2			
Auditory-Reliable	0.36 +- 0.03	0.23 +- 0.03	0.35 +- 0.05
Visual-Reliable	0.18 +- 0.03	0.29 +- 0.03	0.30 +- 0.05
Experiment 3			
Auditory-Reliable	0.27 +- 0.02	0.17 +- 0.02	0.14 +- 0.06
Visual-Reliable	0.2 +- 0.02	0.24 +- 0.02	0.2 +- 0.08
Experiment 4			
Visual undegraded		0.37 +- 0.04	
Visual mid-blur		0.34 +- 0.04	
Visual high-blur		0.32 +- 0.04	

Table 2.3 BOLD amplitudes in all experiments

BOLD amplitudes in auditory cortex, visual cortex and STS (average percent signal change +- SEM).

The average BOLD amplitude in the STS was similar in the two conditions, but there were fluctuations in the amplitude of response to different words. Comparing the STS fluctuations with the fluctuations observed in auditory and visual cortex allows us to measure the functional connectivity of sensory cortex and STS: if the STS weights inputs from early sensory cortex depending on reliability, STS fluctuations might correspond to auditory cortex fluctuations during auditory-reliable words and to visual cortex fluctuations during visual-reliable words. For each subject, a structural equation model was created and tested using the timecourse of the BOLD response to all auditory-reliable and visual-reliable words (Figure 2.3). An ANOVA across subjects on the path coefficients revealed a significant interaction ($F_{(1,9)} = 8.9$, $p = 0.02$) driven by a stronger connection weight between auditory cortex and STS for auditory-reliable words ($p_{LSD} < 0.05$) and between visual cortex and STS for visual-reliable words ($p_{LSD} < 0.05$; Figure 2.4; Table 2.4).

	Aud -> STS	Vis -> STS
Experiment 1		
Auditory-Reliable	0.44 +- 0.03	0.30 +- 0.04
Visual-Reliable	0.33 +- 0.03	0.40 +- 0.02
Experiment 2		
Auditory-Reliable	0.42 +- 0.02	0.26 +- 0.03
Visual-Reliable	0.31 +- 0.02	0.39 +- 0.02
Experiment 3		
Auditory-Reliable	0.50 +- 0.02	0.25 +- 0.03
Visual-Reliable	0.32 +- 0.03	0.40 +- 0.05
Experiment 4		
Visual undegraded		0.50 +- 0.06
Visual mid-blur		0.41 +- 0.06
Visual high-blur		0.32 +- 0.07

Table 2.4 Unidirectional connection weights in all experiments

Connection weights from auditory cortex and visual cortex to STS (average path coefficient +- SEM).

Both the BOLD amplitudes within auditory and visual cortex and their connection weights to STS were modulated by the reliability of the speech stimuli. However, correlating these values across subjects resulted in low correlation values that were not significant (auditory cortex: r = 0.44, p = 0.20 for auditory-reliable and r = 0.14, p = 0.70 for visual-reliable; visual cortex: r = -0.08, p = 0.66 for visual-reliable and r = 0.16, p = 0.85 for auditory-reliable), suggesting that changes in BOLD amplitude and connection weight may be subserved by independent neural mechanisms (Figure 2.5).

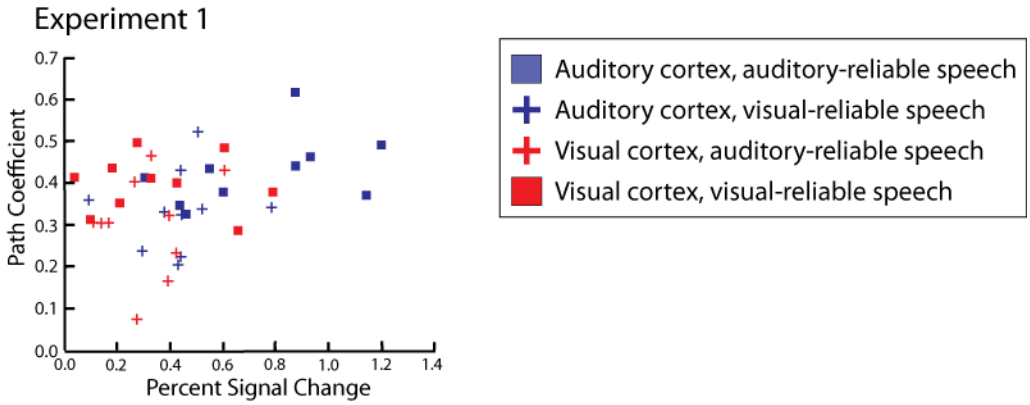

Figure 2.5 Connection weights vs. BOLD amplitudes: Experiment 1

Connection weights vs. BOLD amplitude across conditions in Experiment 1, one symbol per subject.

Our initial analysis measured the connection strength between the STS and ROIs created from independent functional localizers. To determine if 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.

Regions with a stronger correlation with STS during auditory-reliable words were concentrated in and around auditory cortex, while regions with a stronger correlation during visual-reliable words were concentrated in lateral occipital cortex (Figure 2.6; Table 2.2b). These regions largely corresponded to the auditory and visual ROIs generated from the localizer. However, there were additional regions showing differential STS connectivity during auditory and visual-reliable stimulation that were not part of the ROIs. A region of the fusiform gyrus, near the location of the fusiform face area, showed stronger connections with the STS during auditory-reliable words. A region of dorsal occipital cortex, near visual area V3A, showed stronger connections with the STS during visual-reliable words. Interestingly, calcarine cortex (the location of V1) did not show condition-dependent changes in connectivity, nor did portions of Heschl's gyrus (the location of primary auditory cortex).

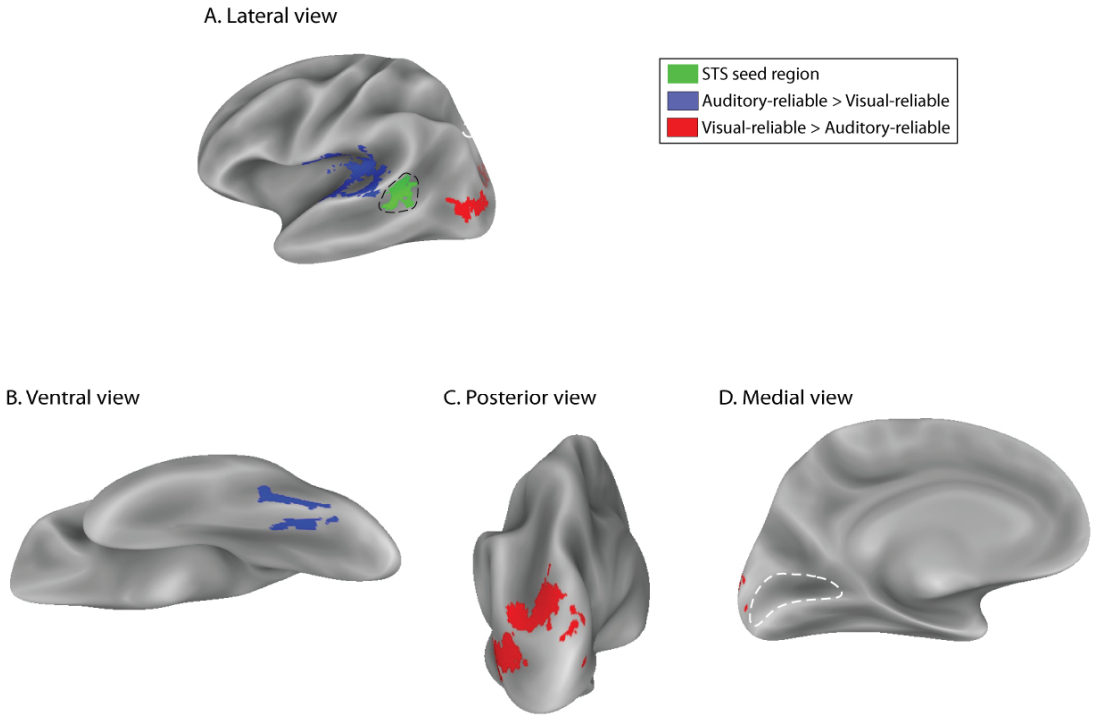

Figure 2.6 Whole-brain connectivity analysis

Whole-brain connectivity analysis showing regions with differential connectivity with STS during auditory-reliable and visual-reliable speech. Group map from ten subjects with STS seed region shown in green surrounded by dashed line. Blue areas showed greater connectivity with the STS during auditory-reliable speech, red areas showed greater connectivity during visual-reliable speech.
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 areas which showed stronger connections with the STS during auditory-reliable words.
C. Dorsal view of the left hemisphere showing a region of dorsal occipital cortex, near visual area V3A, that showed stronger connections with the STS during visual-reliable words.
D. Medial view of the left hemisphere that shows no condition-dependent changes in connectivity in calcarine sulcus, the location of V1, delineated by the white dotted line.

Experiment 2: Rapid-Event Related Experiment

In Experiment 1, we observed stimulus reliability-related changes in BOLD responses and connection weights. However, attention can also increase both the BOLD response and connection weights (92, 96). In Experiment 1, all words within a block were reliable in one modality and unreliable in the other. To prevent subjects from focusing sustained attention on one modality, in Experiment 2, auditory-reliable and visual-reliable words were randomly intermixed using a rapid event-related design.

As in Experiment 1, we predicted an interaction between reliability and sensory cortex, with the sensory cortex processing the reliable stimuli responding more strongly and showing a stronger connection to STS. The ANOVA on the Experiment 2 BOLD amplitudes revealed a significant interaction between ROI and stimulus condition ($F_{(1,9)} = 46.0$, $p = 0.00008$) driven by a greater response to auditory-reliable words in auditory cortex ($p_{LSD} < 0.0001$) and to visual-reliable words in visual cortex ($p_{LSD} < 0.01$). The ANOVA on the Experiment 2 path coefficients also revealed a significant interaction ($F_{(1,9)} = 30.1$, $p = 0.0004$) driven by a stronger connection weight between auditory cortex and STS for auditory-reliable words ($p_{LSD} < 0.01$) and between visual cortex and STS for visual-reliable words ($p_{LSD} < 0.005$). The modality with the greatest effect on STS depended on reliability (Figure 2.7): during auditory-reliable words, auditory cortex had a stronger connection with STS ($p_{LSD} < 0.001$), while during visual-reliable words, visual cortex had a stronger connection with STS ($p_{LSD} < 0.05$).

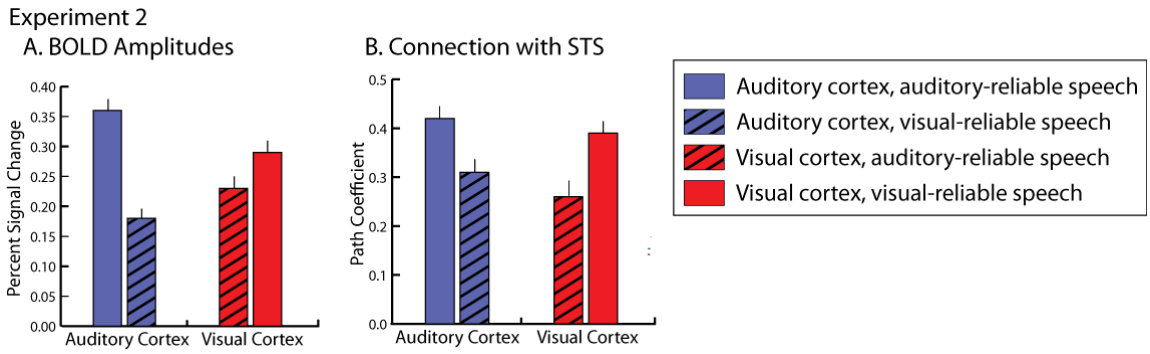

Figure 2.7 BOLD amplitudes and connection weights in Experiment 2

A. BOLD amplitudes in Experiment 2 reported as percent signal change. Error bars are standard error of the mean.
B. Connection weights between auditory and visual cortex and STS in Experiment 2 reported as path coefficients.

As in Experiment 1, no statistically significant correlations were observed between the BOLD amplitude and the connection weight across subjects (auditory cortex: r = -0.06, p = 0.87 for auditory-reliable and r = -0.15, p = 0.68 for visual-reliable; visual cortex: r = -0.10, p = 0.78 for visual-reliable and r = -0.41, 0.25 for auditory-reliable; Figure 2.8).

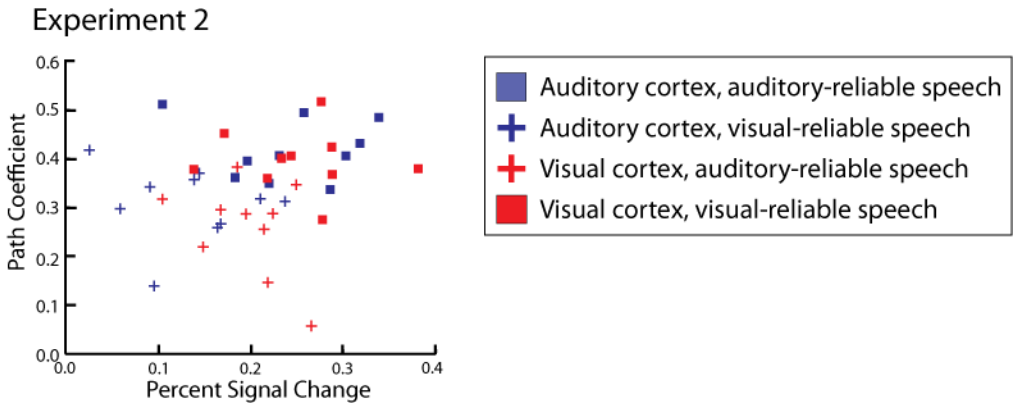

Figure 2.8 Connection weights vs. BOLD amplitudes: Experiment 2

Connection weights vs. BOLD amplitude across conditions in Experiment 1, one symbol per subject.

Additional SEM Models for Experiments 1 and 2

A limitation of structural equation models is their dependence on the initial assumptions about connections within the network. In our simple hierarchical model, auditory and visual cortex both project to STS in a unidirectional fashion; however, most cortical connections are likely to be bidirectional. When we modified the connections in our model to be bidirectional, we observed a similar degree of reliability-weighting (Figure 2.9; Table 2.5). There was a significant interaction between reliability and sensory cortex for the bidirectional path coefficients in both Experiments 1 ($F_{(1,9)} = 13.7$, $p = 0.005$; auditory cortex, $p_{LSD} < 0.01$; visual cortex, $p_{LSD} < 0.05$) and Experiment 2 ($F_{(1,9)} = 24.7$, $p = 0.0008$; auditory cortex, $p_{LSD} < 0.005$; visual cortex, $p_{LSD} < 0.01$).

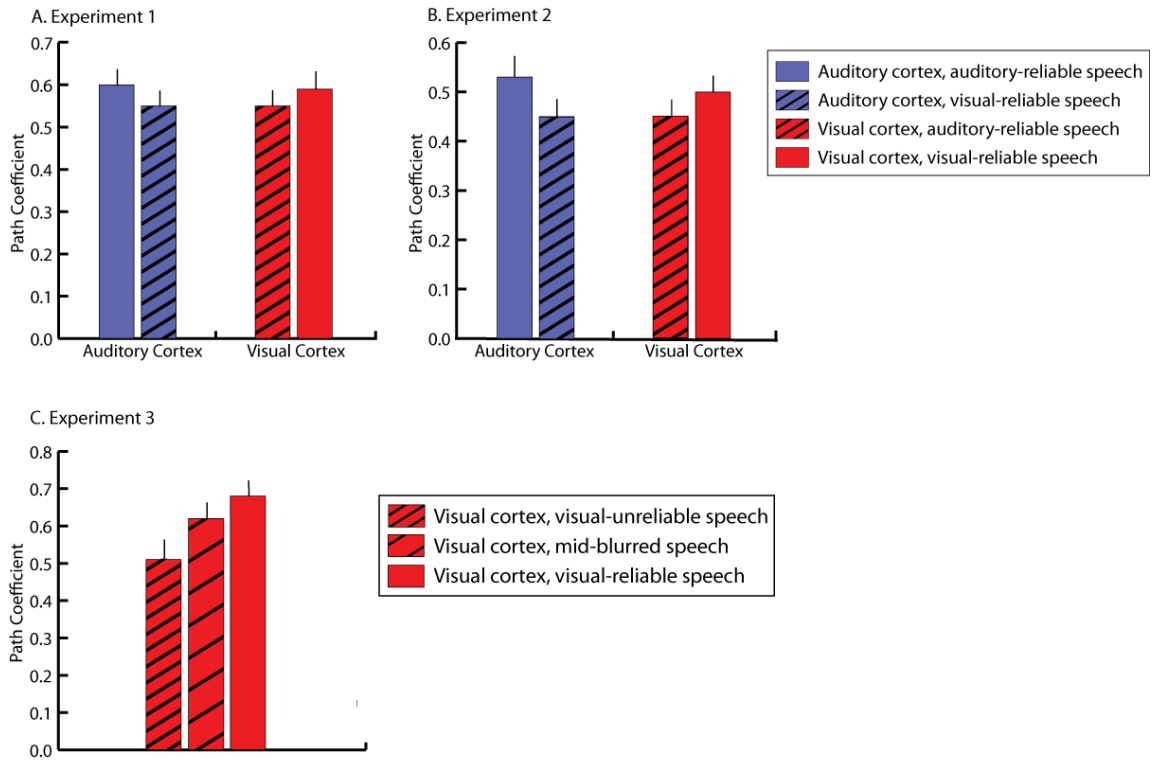

Figure 2.9 Bidirectional connection weights in Experiments 1-3

A. Connection weights between auditory and visual cortex and STS in Experiment 1 reported as path coefficients. Bidirectional connections; compare with results for unidirectional connections in Fig. 2B.
B. Bidirectional weights in Experiment 2.
C. Bidirectional weights in Experiment 3, visual cortex.

	Aud <-> STS	Vis <-> STS
Experiment 1		
Auditory-Reliable	0.60 +- 0.04	0.55 +- 0.05
Visual-Reliable	0.55 +- 0.05	0.59 +- 0.04
Experiment 2		
Auditory-Reliable	0.53 +- 0.04	0.45 +- 0.04
Visual-Reliable	0.45 +- 0.03	0.50 +- 0.03
Experiment 3		
Auditory-Reliable	0.62 +- 0.02	0.46 +- 0.05
Visual-Reliable	0.52 +- 0.04	0.56 +- 0.05
Experiment 4		
Visual undegraded		0.68 +- 0.04
Visual mid-blur		0.62 +- 0.04
Visual high-blur		0.51 +- 0.05

Table 2.5 Bidirectional connection weights in all experiments

A. BOLD amplitudes in auditory cortex, visual cortex and STS (average percent signal change +- SEM).
B. Connection weights from auditory cortex and visual cortex to STS (average path coefficient +- SEM).
C. Bidirectional connection weights between auditory and visual cortex and STS (average path coefficient +- SEM).

We restricted our initial analyses of BOLD activation and connectivity to the left hemisphere because it is the dominant hemisphere for language. An additional analysis was performed to determine if the same pattern of reliability-weighting extended to the right hemisphere (Figure 2.10). The ANOVA on the right hemisphere BOLD amplitudes revealed a significant interaction between ROI and stimulus condition in Experiment 1 ($F_{(1,9)} = 111.6$, $p = 0.000002$; auditory cortex, $p_{LSD} < 0.0005$; visual cortex, $p_{LSD} < 0.001$) and Experiment 2 ($F_{(1,9)} = 31.9$, $p = 0.0003$; auditory cortex, $p_{LSD} < 0.0001$; visual cortex, $p_{LSD} < 0.20$), and the ANOVA on the right hemisphere path coefficients revealed a non-significant interaction between ROI and stimulus condition in Experiment 1 ($F_{(1,9)} = 1.85$, $p = 0.21$) and a significant interaction in Experiment 2 ($F_{(1,9)} = 13.4$, $p = 0.005$; auditory cortex, $p_{LSD} < 0.01$; visual cortex, $p_{LSD} < 0.05$).

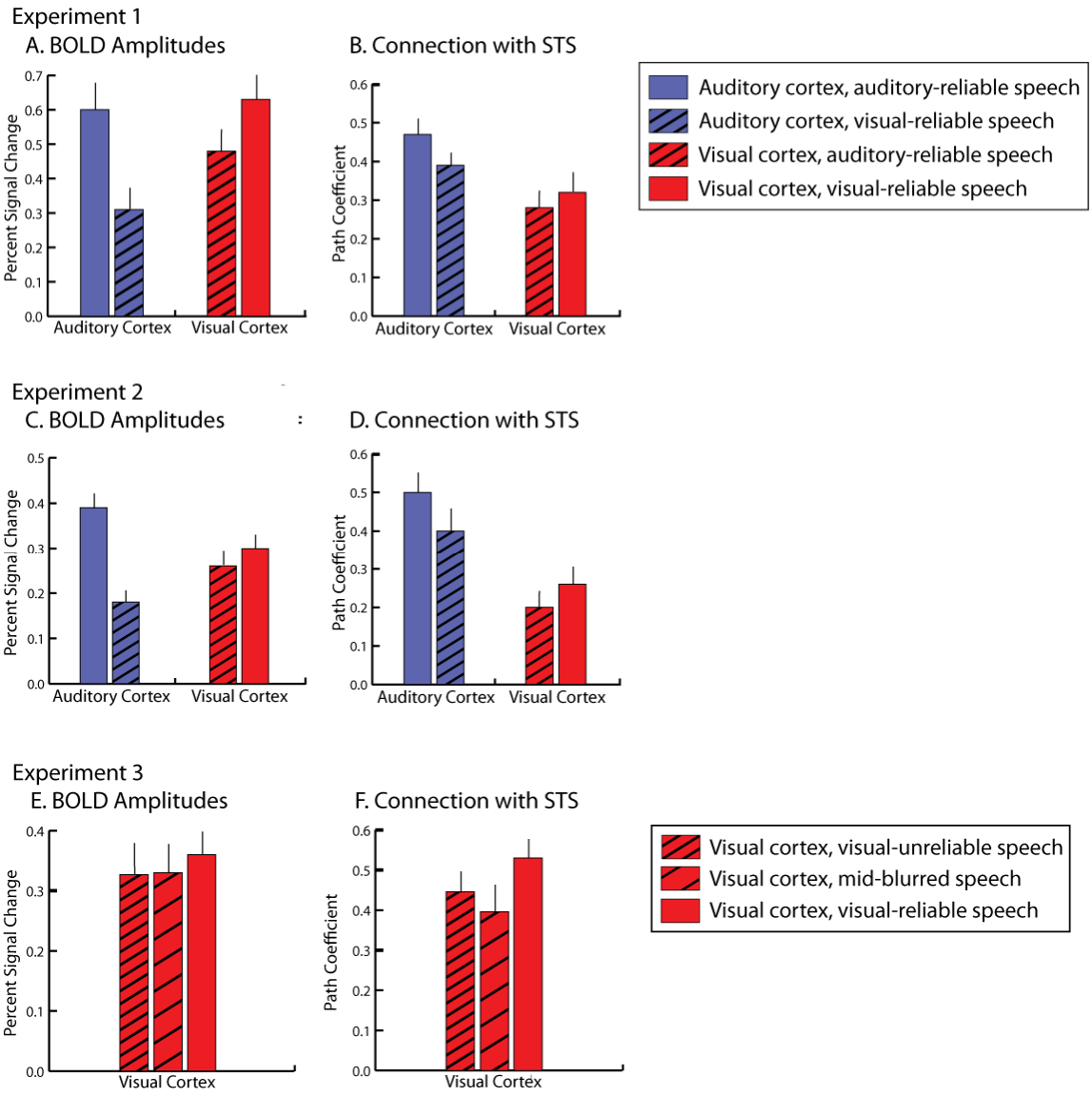

Figure 2.10 Connection weights in right hemisphere analyses

A. BOLD amplitudes in right auditory and visual cortex in Experiment 1 reported as percent signal change.
B. Connection weights between right auditory and visual cortex and STS in Experiment 1 reported as path coefficients.
C. Right hemisphere BOLD amplitudes in Experiment 2.
D. Right hemisphere connection weights in Experiment 2.
E. Right hemisphere visual cortex BOLD amplitudes in Experiment 3.
F. Right hemisphere connection weights between visual cortex and STS in Experiment 3.

Experiment 3: fMRI Rapid Event-Related Parametric Design

In Experiments 1 and 2, the reliabilities of the auditory and visual modalities were not varied independently. This made it impossible to determine if the observed changes in BOLD amplitude and connection weights were driven by auditory reliability, visual reliability or both. Therefore, in Experiment 3, we varied the reliability of the visual modality while holding the reliability of the auditory modality constant. A parametric design with three levels of visual reliability was used in order to determine if BOLD amplitude and connection weights can vary parametrically, as predicted by behavioral models of optimal multisensory integration.

Since three levels of visual reliability were used with a fixed level of auditory reliability, we predicted that reliability-weighting should manifest itself as a main effect of stimulus condition, as opposed to the interactions observed in Experiments 1 and 2. The ANOVA on the BOLD amplitudes in the visual ROI did *not* show a significant main effect of reliability ($F_{(1,5)} = 2.07$, $p = 0.18$), while the ANOVA on the path coefficients *did* show a significant main effect of reliability ($F_{(1,5)} = 17.9$, $p = 0.0005$; visual-reliable vs. visual-unreliable, $p_{LSD} < 0.005$; visual-reliable vs. visual mid-blurred, $p_{LSD} < 0.05$; visual mid-blurred vs. visual-unreliable, $p_{LSD} < 0.05$; Figure 2.11).

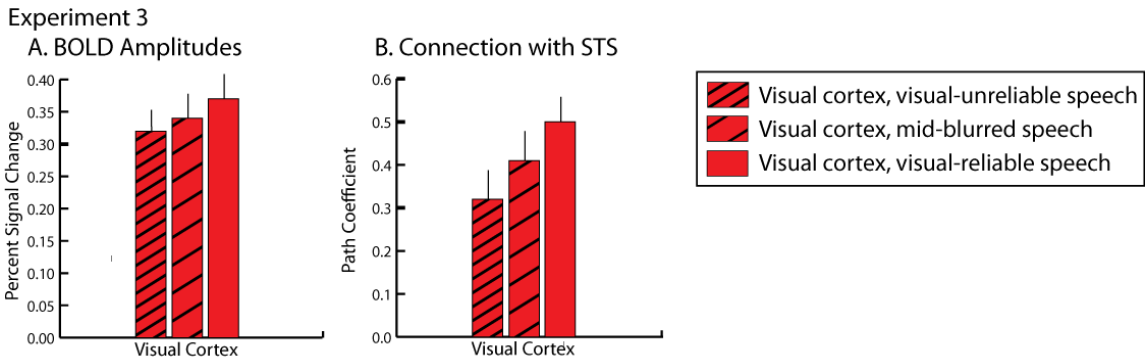

Figure 2.11 BOLD amplitudes and connection weights in Experiment 3

A. BOLD amplitudes in Experiment 3.
B. Connection weights between visual cortex and STS in Experiment 3.

If bidirectional path coefficients were specified in the left hemisphere model, the significant main effect of reliability on connection weights remained ($F_{(1,5)} = 39.1$, $p = 0.000019$; visual-reliable vs. visual-unreliable, $p_{LSD} < 0.0005$; visual-reliable vs. visual mid-blurred, $p_{LSD} < 0.05$; visual mid-blurred vs. visual-unreliable, $p_{LSD} < 0.005$). In the right hemisphere analysis, there was no main effect for either the BOLD amplitudes ($F_{(1,5)} = 0.61$, $p = 0.56$) or the path coefficients ($F_{(1,5)} = 1.85$, $p = 0.21$). No statistically significant correlations were observed between the BOLD amplitude and the connection weight across subjects (visual cortex: $r = -0.42$, $p = 0.41$ with no blurring, $r = 0.18$, $p = 0.73$ with medium blurring, and $r = -0.21$, $p = 0.69$ with high blurring; Figure 2.12).

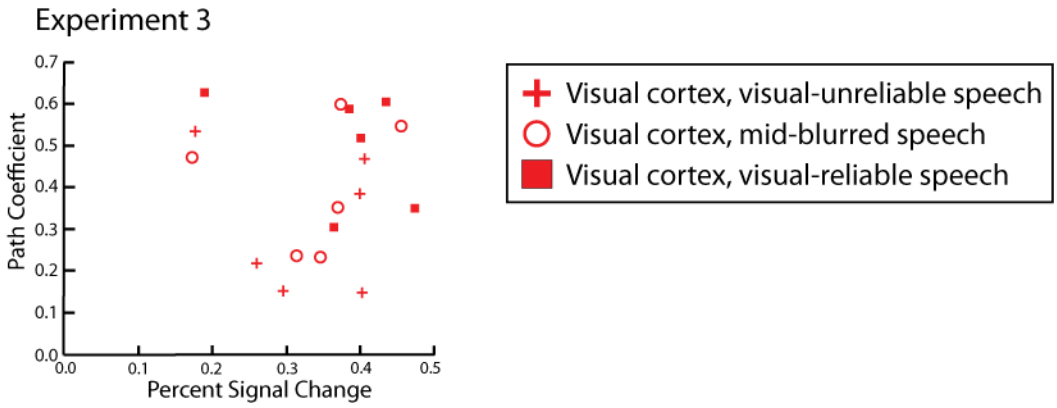

Figure 2.12 Connection weights vs. BOLD amplitudes: Experiment 3

Connection weights vs. BOLD amplitude across conditions in Experiment 3, one symbol per subject.

Experiment 4: Attention Experiment

In Experiment 4, the behavioral task was manipulated to direct subjects' attention to either the auditory or visual modality during presentation of visual-reliable or auditory-reliable syllables. Even when attention was directed away from the reliable modality, there was a significant interaction between sensory cortex and reliability in the same direction as Experiments 1 and 2 for both the BOLD amplitudes ($F_{(1,5)} = 8.7$, $p = 0.03$) and path coefficients ($F_{(1,5)} = 21.9$, $p = 0.005$; Figure 2.13). While there was a significant interaction effect of reliability, there was *not* a significant interaction between cortex and attentional condition, between reliability and attentional condition, or between attention, cortex and reliability. Therefore, it is unlikely that attention is the sole moderator of the observed reliability effects.

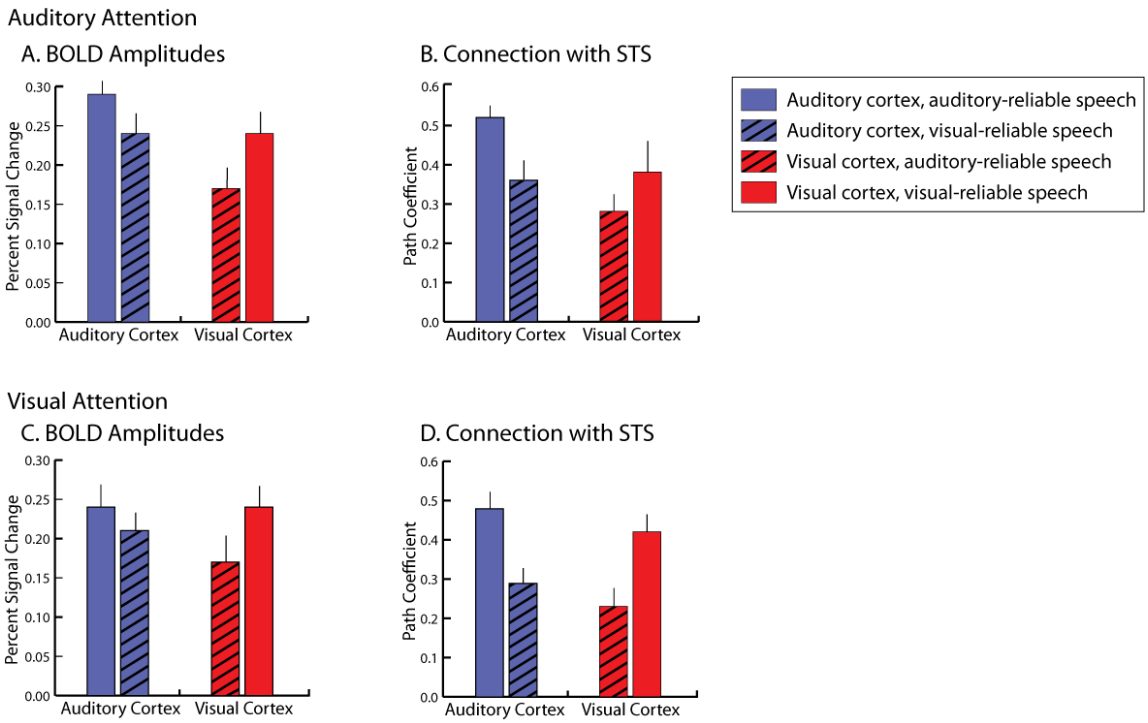

Figure 2.13 BOLD amplitudes and connection weights in Experiment 4

A. BOLD amplitudes during auditory attention reported as percent signal change.
B. Connection weights between auditory and visual cortex and STS during auditory attention reported as path coefficients.
C. BOLD amplitudes during visual attention.
D. Connection weights during visual attention.

Subjects accurately performed the instructed task (97% during auditory attention and 93% during visual attention; chance performance of 50%), indicating that subjects attended to the correct modality. When attention was directed either towards or away from the reliable modality (Figure 2.13), a 3-way ANOVA (with sensory cortex, stimulus reliability and attentional state as factors) on the BOLD amplitudes revealed a significant interaction between sensory cortex and stimulus reliability ($F_{(1,5)} = 8.7$, $p = 0.03$; auditory cortex, $p_{LSD} < 0.20$; visual cortex, $p_{LSD} < 0.10$). Similarly, a 3-way ANOVA on the path coefficients also revealed a significant interaction between cortex and reliability ($F_{(1,5)} = 21.9$, $p = 0.005$; auditory cortex, $p_{LSD} < 0.10$; visual cortex, $p_{LSD} < 0.10$). There was not a significant interaction between attention and cortex (BOLD amplitudes: $F_{(1,5)} = 1.27$, $p = 0.31$; path coefficients: $F_{(1,5)} = 0.52$, $p = 0.48$), between attention and reliability (BOLD amplitudes: $F_{(1,5)} = 0.14$, $p = 0.72$; path coefficients: $F_{(1,5)} = 0.11$, $p = 0.74$) or between attention, cortex and reliability (BOLD amplitudes: $F_{(1,5)} = 0.41$, $p = 0.55$; path coefficients: $F_{(1,5)} = 0.76$, $p = 0.39$). The fact that we observed a significant effect of reliability but not of attention suggests that attention is not the sole moderator of these effects.

Experiment 5: Behavioral Experiment

In order to replicate previous studies demonstrating that perception of audiovisual speech is driven by the more reliable modality, in Experiment 5 we created single syllables that were reliable in either the auditory or visual modality. 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. This effect was observed in each of ten

subjects (p = 0.0001, paired t-test; Figure 2.14).

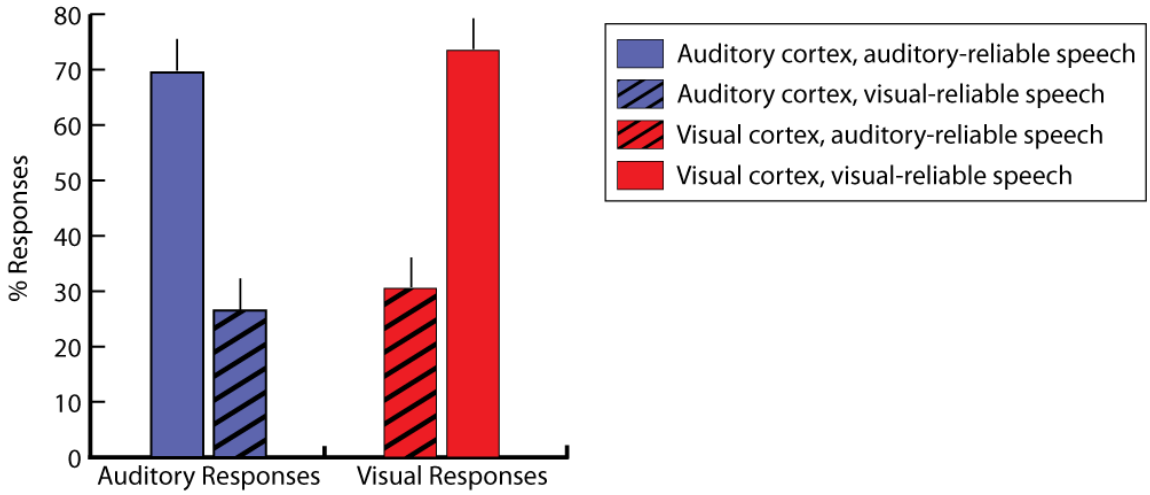

Figure 2.14 Reliability-weighted perception in Experiment 5

Subjects' perception of incongruent audiovisual syllables (Experiment 5).

Conclusions

The auditory and visual modalities both play a role in human speech perception (1, 97-100), making speech an important example of multisensory integration. As we observed in experiment 5, perception of audiovisual speech is driven by the more reliable modality, regardless of whether it is auditory or visual (10, 20, 101). To understand the neural mechanisms for perceptual reliability-weighting, we conducted behavioral and fMRI experiments in which subjects were presented with audiovisual speech of varying reliability. More reliable stimuli evoked a stronger BOLD response in sensory cortex and resulted in a stronger connection weight between the sensory cortex representing the reliable stimulus modality and the STS. The change in connection weights was striking: the dominant modality, defined as the sensory modality with the strongest input to STS, was determined by reliability. We propose a simple model of reliability-weighted speech perception. First, stimuli of differing reliability evoke distinct responses in sensory cortex. Second, the STS weights the responses from each sensory cortex by that modality's reliability. This, in turn, produces perceptual reliability-weighting.

We adjusted reliability by degrading our auditory and visual stimuli (7, 8). Auditory neurons have sharp peaks in frequency space (102, 103), and blurring the spectral information reduces single-unit responses (104). Therefore, auditory speech degraded using a noise-vocoded filter (as used in our study) results in a reduced BOLD response in auditory cortex relative to undegraded speech (105, 106). Visual neurons respond to high-contrast edges (107, 108), and low contrast edges result in weaker neural responses in visual cortex (109, 110). Therefore, low contrast images (such as the blurred videos in our study) result in reduced fMRI activity in visual cortex (111, 112).

In sum, the changes in BOLD amplitude in sensory cortex for our different stimuli can be most parsimoniously explained as reflecting low-level stimulus properties.

In the second stage of the model, activity in sensory cortex is integrated by the STS with weights dependent on the reliability of each modality. Interrupting activity in the STS modifies perception of audiovisual speech in humans (113), supporting a role for the STS in auditory-visual integration of speech (42, 114). In four separate fMRI experiments, we observed a consistent pattern of STS connection weights: the connection weight from the more reliable sensory cortex to the STS was stronger than the connection weight from the less reliable sensory cortex to the STS. This reliability-dominated input into STS could serves as the neural basis for the behavioral observation that speech perception is driven by the more reliable modality, regardless of whether it is auditory or visual (10, 20, 101).

In our model, the sensory cortex responses evoked by unreliable stimuli (step 1) are distinct from the integration of those responses by the STS (step 2). Across subjects and experiments, the changes in BOLD amplitude in sensory cortex were uncorrelated with the changes in connection weights between sensory cortex and STS, supporting a two-step model. Additional evidence comes from a recent study of visual-tactile integration in which the stimulus was made less reliable by adding dynamic noise (instead of filtering, as in the present study) (115). Adding dynamic noise resulted in *increased* BOLD amplitude in sensory cortex for unreliable stimuli (as opposed to the *decreased* BOLD amplitude for unreliable stimuli in the present study). Despite the opposite patterns of BOLD amplitudes, in both studies the connection weights between sensory cortex and multisensory cortex were reliability-weighted, with stronger

connections between the sensory cortex representing the reliable stimulus modality and the multisensory area. In Experiment 3 of the present study, we observed significant changes in connection weights but not in BOLD amplitudes as reliability was parametrically varied, also suggesting that changes in connection weights may be more important than BOLD amplitude changes for perceptual reliability-weighting.

For our initial analysis, we chose a structural equation model 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 (116-120). When incorporating bidirectional connections into the structural equation model, we also observed robust reliability-weighting, confirming that reliability-weighted connections are consistent across different models. The whole-brain connectivity analysis also showed enhanced connectivity between auditory cortex and visual cortex and STS for more reliable stimulation. 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 (46, 121, 122). These connections (rather than connections between association and primary areas) may be most important for reliability-weighted speech perception. 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 (123, 124), 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.

How could the STS compute reliability in order to properly assign the connection weights to each modality? The simplest model is that the sensory cortex itself assesses the reliability of the stimuli in its modality and adjusts the synaptic weights of its projections to STS proportionally. A number of cellular mechanisms could underlie the changes in synaptic weights, such as spike timing-dependent plasticity (125). The assessment of reliability could also be performed in a number of ways. One possibility is simply that the summed activity in the sensory cortex indicates the level of reliability. However, this explanation is unlikely, as visual cortex did not show greater activity for reliable than unreliable stimuli. Another possibility is that the STS performs a separate computation on the reliability of an input modality, independent of the amplitude of the response. Using this information, the STS could upregulate or downregulate the synaptic weights of the pathways carrying that information. A possible candidate for this computation is the "sharpness" of the population response, as posited for normalization models of attention that divide the strongest response in the input population by the pooled background activity (126, 127). A strong peak for neurons responding to a particular stimulus (*e.g.*, auditory "pen" or "road") indicates that a great deal of unambiguous information about stimulus identity is available from that modality, suggesting that it is reliable and should be given a high weight. Conversely, a low selectivity peak (*e.g.*, similar responses for pools of neurons responding to "white" or "write") suggests that there is relatively little unambiguous information about stimulus identity available in that modality and that it should be given low weight. Our results can

be also be interpreted in light of predictive coding models of cortical function (128). 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 (129), 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.

Computational models have suggested that reliability-weighting could occur by a simple linear summation of neuronal responses (49, 130) that are stronger during reliable stimuli and weaker during unreliable stimuli. However, an explicit prediction of these models is that connection weights between areas do *not* change depending on the reliability of the stimulus. In each of our experiments, we observed a significant change in the connection weights driven by reliability, as did a recent fMRI study of visual-tactile integration (115) and recent electrophysiological studies of visual-vestibular multisensory integration in macaque monkeys (131, 132).

In order to form a coherent audiovisual percept during presentation of speech, multisensory brain areas must combine information from both the auditory and visual cortex. A popular idea for how this may occur is through oscillations and synchrony (133, 134). For example, if auditory and visual neurons are firing in phase, their corresponding percepts will more likely be fused.

Temporal synchrony between firing of sensory areas and STS may mediate perception of audiovisual speech with reliable and unreliable components. If stimuli are

more reliable in one modality than the other, this may create stronger oscillations within a sensory cortex that would entrain downstream areas. For instance, STS would fire synchronously with auditory cortex during auditory-reliable stimuli. This synchronous firing would allow the sensory area to elicit activity in downstream areas that are responsible for multisensory perception, and thus would create a percept more similar to auditory stimulus than the visual stimulus. The reverse would be true for visual-reliable stimuli: visual cortex would fire synchronously with STS and drive the percept towards the visual stimulus. Of course, we cannot directly observe neural synchrony with fMRI. However, computational models suggest that effective connectivity as measured with neuroimaging increases with synchronous firing between areas (135).

Behavioral studies have shown that when one modality contains more reliable information than the other in an audiovisual stimulus, perception tends to follow the rules of optimal integration. In the case of audiovisual speech perception, optimal integration predicts that the multisensory estimate of an audiovisual word is between the estimates from auditory and visual information but closer to the estimate of the more reliable modality. Witten and Knudsen demonstrated that the ventriloquist effect is an example of optimal integration, in which perception more closely corresponds to reliable visual information when the auditory information is less reliable (9). Similarly, Ma et al. showed that low auditory reliability increased reports of the visual word while high auditory reliability increased reports of the auditory word (10). We found evidence for behavioral reliability weighting using incongruent audiovisual speech stimuli; subjects were more likely to report perception of the auditory syllable during auditory-reliable stimuli and were more likely to report perception of the visual syllable during visual-

reliable stimuli. Our behavioral results are consistent with the idea of optimal integration by showing increased responses corresponding to reliable modality.

In addition to the finding of optimal integration, behavioral studies have also shown that reliability-weighting occurs even if subjects are forced to attend to one modality, suggesting that reliability-weighting is independent of modality-specific attention (136). 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 either 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.

Many fMRI studies of audiovisual speech perception employ a task in order to ensure proper attention as well as to monitor behavioral perception during the experiment. One concern with the study of BOLD activation during active tasks, however, is the role of task on hemodynamic response. For instance, van Atteveldt et al. (2007) found that the STS responded less to incongruent audiovisual stimuli than congruent stimuli during passive presentation (137). However, when subjects made a decision about whether or not the stimulus was congruent, this difference was abolished. In our fMRI studies, STS activity was not significantly different during auditory-reliable and visual-reliable stimuli, whether or not there was a task. Additionally, our finding of increased connectivity between early sensory areas processing reliable stimuli and

STSms was consistent both with passive viewing of stimuli and with a 2AFC task during each stimulus.

Previous studies have demonstrated that connection weights between sensory cortex and higher areas can vary depending on the behavioral context and the stimuli presented to the subject (138-146), 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) investigated cortical activation and connectivity during temporally congruent streams of auditory tones and visual patterns as compared with temporally incongruent audiovisual stimuli and unisensory auditory and visual stimuli. They found that activation in auditory cortex, visual cortex and multisensory STS was elevated during congruent audiovisual stimuli compared with incongruent audiovisual stimuli. Noppeney et al. (2007) observed increased connection strengths from auditory cortex to STS during auditory speech when paired with an incongruent visual word, suggesting that strengthened connection from sensory to multisensory areas may aid in understanding out-of-context speech. In Patel et al. (2006), subjects listened to sentences that were either different in content (different sentences) or different in speaker. The authors found a stronger connection from Wernicke's area to the superior temporal gyrus and the posterior cingulate gyrus while passively listening to different sentences rather than the same sentence repeatedly. Husain et al. (2006) studied cortical activity during speech and non-speech sounds and found stronger functional connectivity between left IFG and auditory cortex during a categorization task than during an

auditory discrimination task.

In this study of reliability-weighting, there were stronger functional connections between STS and cortical areas that process the more reliable modality presented. This pattern of connectivity is sensible from the standpoint of optimal multisensory integration. 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. This is the effect we observed in our fMRI experiments, and it mirrors the weighting that has been observed in behavioral studies, in which the more reliable modality has greater influence on the behavioral decision (7-10). In summary, these fMRI results suggest that strengthened STS functional connectivity may provide a general mechanism for heightened multisensory integration.

CHAPTER 3: STS ACTIVITY CORRELATION WITH MCGURK PERCEPTION

Introduction

Understanding speech is an inherently multisensory task; independent information available from the auditory modality (heard speech) and the visual modality (mouth movements) are combined under everyday conditions. These visual cues generally improve comprehension, especially in noisy environments (2, 3, 11). However, visual input from mouth movements can be so compelling as to change perception of clear auditory speech. A remarkable illusion known as the McGurk effect (11) is a powerful demonstration of this process: an auditory "ba" presented with the mouth movements of "ga" is perceived by the listener as a completely different syllable, "da" (referred to as the McGurk percept).

However, the McGurk effect is not experienced by all subjects, with population estimates of McGurk susceptibility ranging from as high as 98% in the original report to as low as 26% (147). Other illusions that require the integration of information across modalities, such as the size-weight illusion, also show substantial inter-subject variation, but little is known about the neural mechanisms for individual differences in susceptibility to any illusion.

The human posterior superior temporal sulcus (STS) is a brain region important for integrating auditory and visual information about both speech and non-speech stimuli studies (20, 39-44, 148). We recently demonstrated that interrupting activity in the STS with transcranial magnetic stimulation (TMS) reduced the frequency of the McGurk effect in subjects who are susceptible to the illusion (47). Instead of the McGurk percept, TMS caused these subjects to perceive only the auditory syllable, the same percept experience by McGurk-resistant individuals.

Since interfering with activity in the STS makes McGurk-susceptible individuals more similar to McGurk-resistant individuals, we hypothesized that differences in STS activity might explain intersubject differences in McGurk susceptibility. To test this hypothesis, we used BOLD fMRI to measure activity in the STS as subjects were presented with congruent and incongruent syllables. Since enhanced neural activity is a signature of the multisensory integration required for the McGurk percept, we predicted that a greater STS response would be observed in McGurk-susceptible individuals than in McGurk-resistant individuals.

To measure STS activity, we used independent localizers to identify the location of the STS multisensory area in each subject. Previous fMRI studies of the McGurk effect did not use functional localizers, which may explain why previous studies did not report, or did not examine, a link between STS activity and McGurk susceptibility (20, 149-152). Without an independent localizer, comparisons are typically performed on a voxel-by-voxel basis in standard space. This makes it difficult to obtain sufficient statistical power, because the number of brain voxels (tens of thousands) is much greater than the number of subjects in neuroimaging studies (10 – 20 in previous fMRI McGurk studies). In addition, the location of the STS multisensory area in standard space varies greatly from subject-to-subject, hindering the ability of voxel-wise analyses to detect a correlation between activity in individual voxels and behavior. The use of functional localizers to identify the STS circumvents both of these difficulties, and ensures statistical independence, a problem that has plagued neuroimaging studies of intersubject differences.

Methods

Subjects and Stimuli

14 healthy right-handed subjects (6 female, mean age 26.1) 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.

The stimulus consisted of a digital video recording of a female speaker speaking "ba", "ga", "da" and "ma" (11). Digital video editing software (iMovie, Apple Computer) was used to modify the original recordings. The duration of the auditory syllables ranged from 0.4 to 0.5 seconds. The total length of each video clip ranged from 1.7 to 1.8 seconds in order to start and end each video in a neutral, mouth-closed position and to include all mouth movements from mouth opening to closing.

Not all incongruent auditory-visual stimuli produce a McGurk percept, defined as a percept not present in the original stimulus. For instance, auditory "ba" + visual "ga" produces the McGurk fused percept of "da", while auditory "ga" + visual "ba" produces an auditory percept such as "ga" or a combination percept such as "g-ba" (11). This non-McGurk incongruent syllable (auditory "ga" + visual "ba") will be referred to as "incongruent" in this manuscript.

Behavioral Pre-Testing

Prior to scanning, each subject's perception of McGurk and incongruent syllables was assessed. Each subject was presented with 10 trials of McGurk syllables (auditory "ba" + visual "ga") and 10 trials of incongruent syllables that do not produce a McGurk percept (auditory "ga" + visual "ba"). Auditory stimuli were delivered through

headphones at approximately 70 dB, and visual stimuli were presented on a computer screen. Subjects were instructed watch the mouth movements and listen to the speaker.

In order to assess perception, subjects were asked to repeat aloud the perceived syllable, with no constraints placed on potential responses: all responses were recorded exactly as spoken. This open-choice response has been shown to be a conservative measure of McGurk perception in previous studies that have compared it with a forced-choice procedure (153, 154) and is more informative with respect to possible intersubject differences in perception. For the McGurk syllables, fused percepts such as "da," "fa" and "va" were used as indicators that subjects perceived the McGurk effect, because they were not present in the original stimulus (11). Responses corresponding to "ba," the auditory stimulus, indicated that subjects did not perceive the McGurk effect.

fMRI Syllables Experiment

Each subject was presented with 3-4 scan series each containing 55 McGurk syllables (2 s each), 55 incongruent syllables (2 s each), 10 target trials (audiovisual "ma") and 30 null trials (2 s of fixation baseline) presented pseudo-randomly in optimal rapid event-related order (95). For 9 subjects, congruent syllables were presented in addition to the McGurk and incongruent syllables. For these subjects, each scan series contained 25 congruent "ba" syllables (2 s each), 25 congruent "ga" syllables (2 s each), 25 McGurk syllables (2 s each), 25 incongruent syllables (2 s each), 10 target trials (audiovisual "ma") and 30 null trials (2 s of fixation baseline) presented pseudo-randomly in optimal rapid event-related order. Each stimulus lasted approximately 1.7-1.8 seconds, with fixation crosshairs occupying the remainder of each 2-second trial. 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. We did not have subjects perform a behavioral task in the scanner during these stimulus conditions to avoid introducing additional task-related activations. Instead, all subjects were instructed to press a button during each target trial.

General fMRI Methods

Anatomical scans for each subject consisted of two T1-weighted scans anatomical collected at 3T using an 8-channel head gradient coil. The two anatomical scans were aligned, averaged into one dataset, transformed to the Talairach coordinate system (68). Each anatomical dataset was normalized to the the N27 reference anatomical volume (69) for group analysis. A three-dimensional cortical surface model was created from these T1-weighted scans using FreeSurfer (70, 71), and functional data was overlaid onto this surface model using SUMA (72).

Functional scans consisted of T2*-weighted images collected using gradient-echo echo-planar imaging (TR = 2015 ms, TE = 30 ms, flip angle = 90°) with in-plane resolution of 2.75 x 2.75 mm. Thirty-three axial slices were collected at 3 mm intervals in order to collect data from the entire cerebral cortex. Each functional scan series consisted of 153 brain volumes. The first three volumes of each scan were discarded, resulting in 150 usable volumes.

Stimuli were presented using Presentation version 12 (Neurobehavioral Systems, Albany, CA). Auditory stimuli were presented to subjects within the scanner using MRI-compatible pneumatic headphones. Visual stimuli were projected onto a screen and subsequently viewed through a mirror attached to the head coil. Button presses were used to assess subject performance of tasks and were collected using a fiber-optic button

response pad (Current Designs, Haverford, PA). An eye tracking system to ensure alertness and visual fixation during all functional scans (Applied Science Laboratories, Bedford, MA).

Analysis of the fMRI data was performed using Analysis of Functional NeuroImages software (AFNI) (73). The false discovery rate (FDR) procedure (74) was used to correct for multiple comparisons, and the FDR's were reported as "q" values. Functional activation was analyzed first within each individual subject, and then data was combined across subjects using a random-effects model. Functional activation maps were aligned to each subject's averaged anatomical scan and were 3-dimensionally motion-corrected using a local Pearson correlation (75). For voxel-wise group analyses, we used a multiple linear regression technique using the AFNI function *3dRegAna* to identify voxels with a significant correlation between activity during McGurk stimuli and McGurk susceptibility.

A deconvolution analysis was performed for each subject to create functional activation maps using the AFNI function *3dDeconvolve*. One regressor was created for each stimulus type and then a convolution was performed to estimate the amplitude of response to each stimulus condition. To help correct for head motion, six movement regressors were created for each scan and were modeled as regressors of no interest.

We performed connectivity analyses to determine if changes in functional connectivity between language areas were correlated with McGurk susceptibility. A structural equation model was constructed and tested for each subject. The model consisted of the four ROIs (auditory cortex, visual cortex, frontal cortex and STS) in the left hemisphere with bidirectional connections between auditory cortex and STS,

between visual cortex and STS and frontal cortex and STS. The amplitude of the hemodynamic response was estimated for each individual McGurk stimulus and averaged within each ROI to produce a vector of 75-100 McGurk amplitudes. These amplitudes were used to calculate the correlation matrix and path coefficients in each subject using the AFNI functions *1ddot* and *1dsem*. The path coefficients obtained from each subject were correlated with each subject's McGurk susceptibility.

fMRI Functional Localizer and Regions of Interest

A key point in our analysis is that the STS ROI was created in completely separate scan series using different stimuli than were used in the McGurk test. It would be trivial indeed if we identified voxels that were correlated the behavioral percept and then averaged only those voxels (155). A functional localizer consisting of blocks of auditory and visual words was used to identify four regions of interest (ROIs) in each subject important for speech processing: auditory cortex, visual cortex, inferior frontal cortex and STS. The ROIs were obtained from separate scan series, apart from the scan series for collecting audiovisual data, in order to prevent bias and avoid the phenomenon of "double-dipping" (77). The ROIs were created only in the left hemisphere because the left hemisphere is dominant for language (78, 79) and were generated separately for each individual because of a high degree of intersubject variability (156).

The functional localizer contained five unisensory auditory and five unisensory visual blocks presented in random order. Each block ten trials (2 seconds each), one undegraded word per trial, and there were 10 seconds of fixation between each block. The auditory, visual, frontal and STS ROIs were created separately for each subject on the cortical surface. Voxels within the STS ROI were chosen within the anatomically-

defined posterior STS for each subject (89, 90). Voxels with activity greater than baseline during both auditory-only and visual-only blocks were used for further analysis ($q < 0.05$ for each modality). Voxels within the auditory ROI were chosen to center on Heschl's gyrus within boundaries for the primary auditory cortex based on prior work (80, 81). These boundaries consisted of the superior temporal gyrus in the lateral direction, the medial termination of Heschl's gyrus in the medial direction, the first temporal sulcus in the anterior direction and the transverse temporal sulcus in the posterior direction. Within these boundaries, voxels with activation greater than baseline during auditory-only blocks were used for further analysis. Voxels within the visual ROI were chosen to center within extrastriate lateral occipital cortex, a brain region critical for processing moving and biological stimuli which includes the middle temporal visual area and the extrastriate body area (82-87). Voxels with along the inferior temporal sulcus (ITS) or its posterior continuation near areas LO and MT (88). Within these boundaries, voxels with activation greater than baseline during visual-only blocks were used for further analysis. The frontal ROI was defined using a conjunction analysis to find all voxels that responded to both auditory and visual words greater than baseline that were located within the anatomically-defined opercular region of the inferior frontal gyrus as well as the inferior portion of the precentral sulcus using an automated parcellation method (89, 157).

Results

Behavioral Testing

In the behavioral pre-test immediately before the MRI experiment, there was a high degree of intersubject variability in McGurk susceptibility (Figure 3.1), ranging from 0% of the McGurk syllables (auditory "ba" + visual "ga) perceived with a fused McGurk percept of "da" (subjects 1-3) to 100% of the McGurk syllables perceived with a McGurk percept (subjects 13-14). The mean percentage across subjects was 46 +- 40%. For incongruent (non-McGurk) syllables consisting of auditory "ga" + visual "ba", none of the subjects experienced a fused "da" percept during the non-McGurk incongruent syllables.

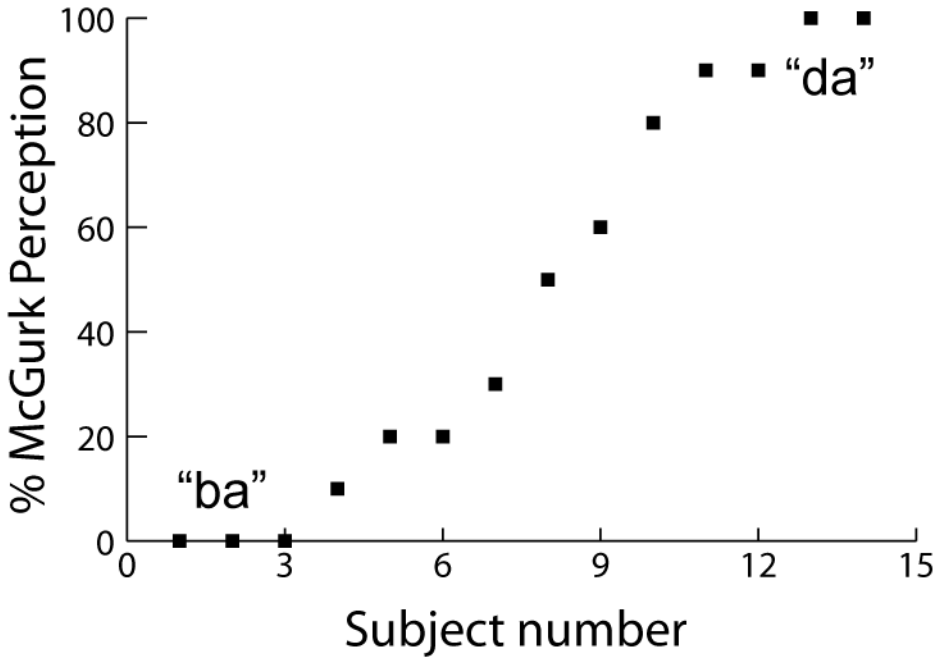

Figure 3.1 McGurk susceptibility across subjects

McGurk susceptibility for each of 14 subjects expressed as a percentage of responses corresponding to the McGurk percept during presentation of McGurk stimuli.

Based on their perception of the McGurk stimuli, we classified subjects into three groups: non-perceivers (6 subjects, susceptibility 0 – 20%), perceivers (5 subjects, susceptibility 80% - 100%) and intermediate perceivers (3 subjects, susceptibility 21% - 79%). To ensure that McGurk susceptibility was stable within subjects, 4 subjects were tested both immediately before and immediately after scanning. The McGurk susceptibility was similar, with a mean difference in pre- and post-test scores of 5% +- 6.5%. None of the subjects shifted groups based on their pre and post-test scores.

fMRI Localizer Experiment

The word stimuli presented in the functional localizer scan series evoked robust hemodynamic responses in auditory cortex for auditory speech and in visual cortex for visual speech. The STS responded strongly to both auditory and visual speech (Figure 3.2). The functional localizers were collected in separate scan series, independent from the experimental scan series described below, and used a completely different stimulus set (without any McGurk stimuli), allowing statistical tests to be performed without bias.

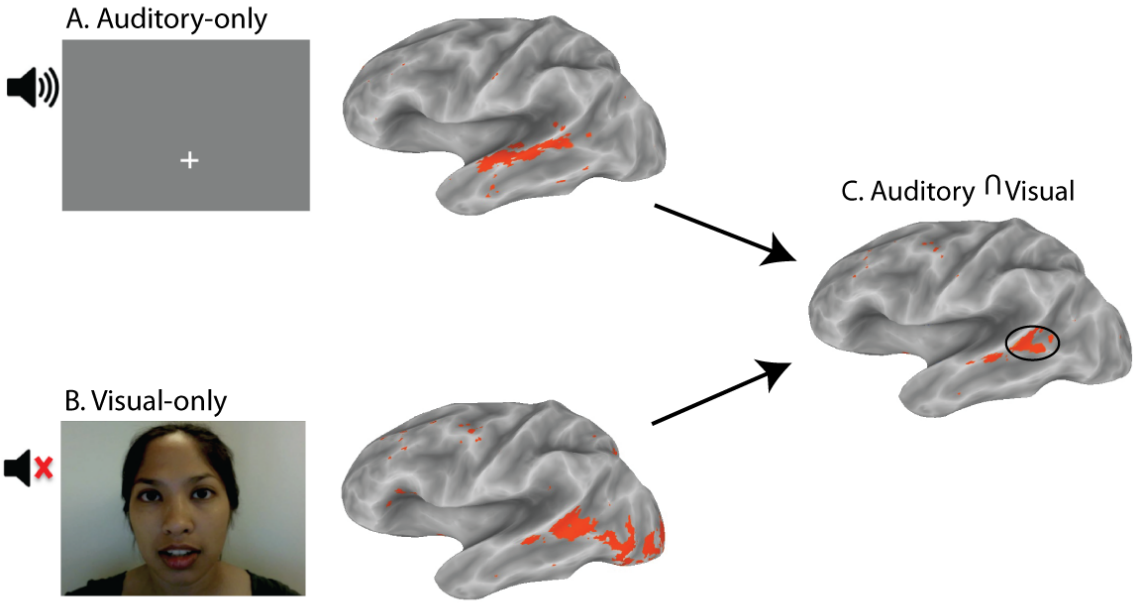

Figure 3.2 Identification of audiovisual areas of STS

A. Undegraded auditory speech (loudspeaker icon) with visual fixation crosshairs. Adjacent cortical surface shows activity in orange during blocks of auditory-only speech.
B. Undegraded visual speech (illustrated by a single frame from a video) with no auditory stimulus. Adjacent cortical surface shows activity during blocks of visual-only speech.
C. Cortical surface shows areas that are active during both auditory-only and visual-only speech blocks.

fMRI McGurk Experiment

Using a rapid-event related design, we measured the brain response to presentation of McGurk syllables, incongruent syllables, and congruent syllables. Our initial analysis focused on our *a priori* region of interest, the left STS (Figure 3.3). Using our categorization of subjects into non-perceivers, perceivers and intermediate perceivers, we did an ANOVA on the STS response (it should be emphasized that the division into groups was completely independent of the STS response, so this analysis was unbiased). There was a significant effect of McGurk susceptibility group on STS response to McGurk syllables ($F_{(2,13)} = 5.2$, $p = 0.03$). The highest perceivers had the highest mean STS response (0.21%+- 0.02), the non-perceivers had the lowest mean STS response (0.10%+- 0.02, significantly less than high, $p = 0.007$), and the intermediate perceivers were closer to the high perceivers (0.18%+- 0.05%, not significantly greater than low perceivers, $p = 0.13$).

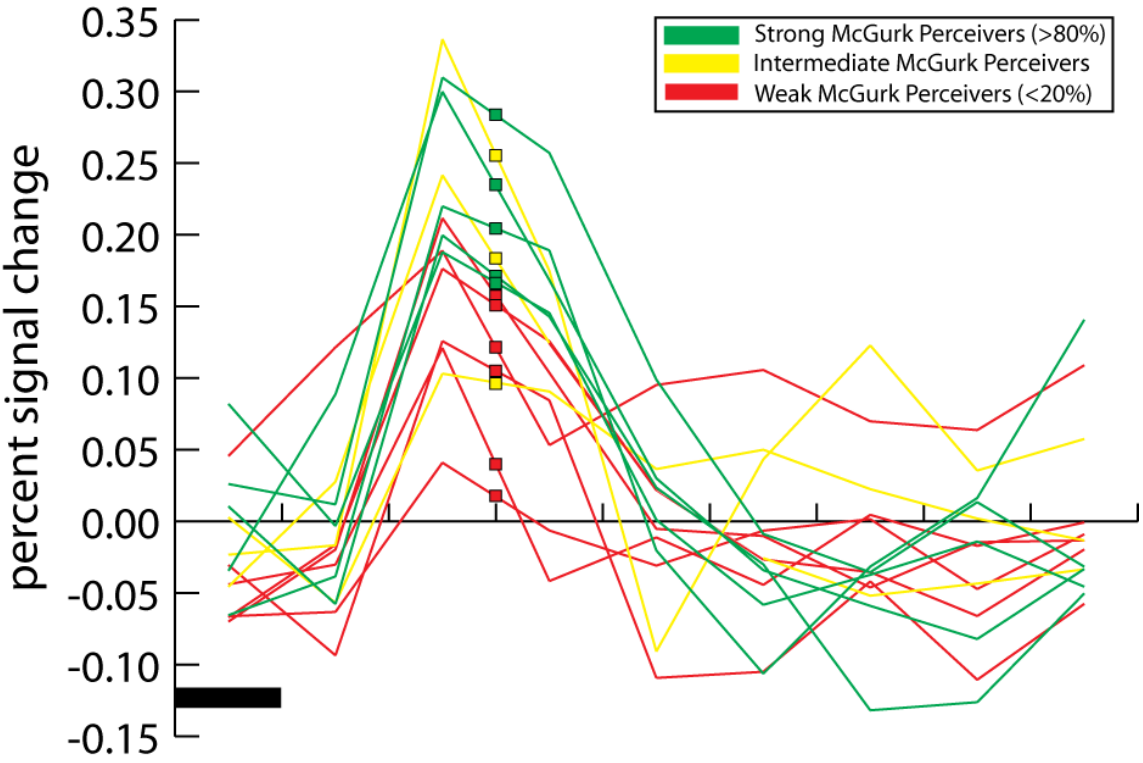

Figure 3.3 STS responses during McGurk stimuli

Each square corresponds to the amplitude of response to McGurk stimuli in an individual subject's STS ROI, defined as the mean response between 4 seconds and 6 seconds after stimulus onset. The green, brown and red tracings represent the average hemodynamic response curves across strong perceivers, intermediate perceivers and non-perceivers, respectively. The black bar represents the time of stimulus onset (0 seconds).

Next, we examined each individual's STS response to McGurk stimuli (Figure 3.4). The subject with the weakest STS response to McGurk syllables (0.02%) had the smallest likelihood of experiencing a McGurk percept (0%); the subject with the strongest STS response (0.28%) had the highest likelihood (100%). Across all subjects, there was a significant positive correlation between each subject's STS response to McGurk syllables and their likelihood of experiencing the McGurk percept ($r = 0.73$, $p = 0.003$). Even excluding the two subjects with the weakest and strongest STS response, the correlation was still significant ($r = 0.63$, $p = 0.03$). A significant correlation was also observed between STS responses to incongruent syllables and McGurk susceptibility, although weaker than the correlation between McGurk syllables ($r = 0.63$, $p = 0.02$).

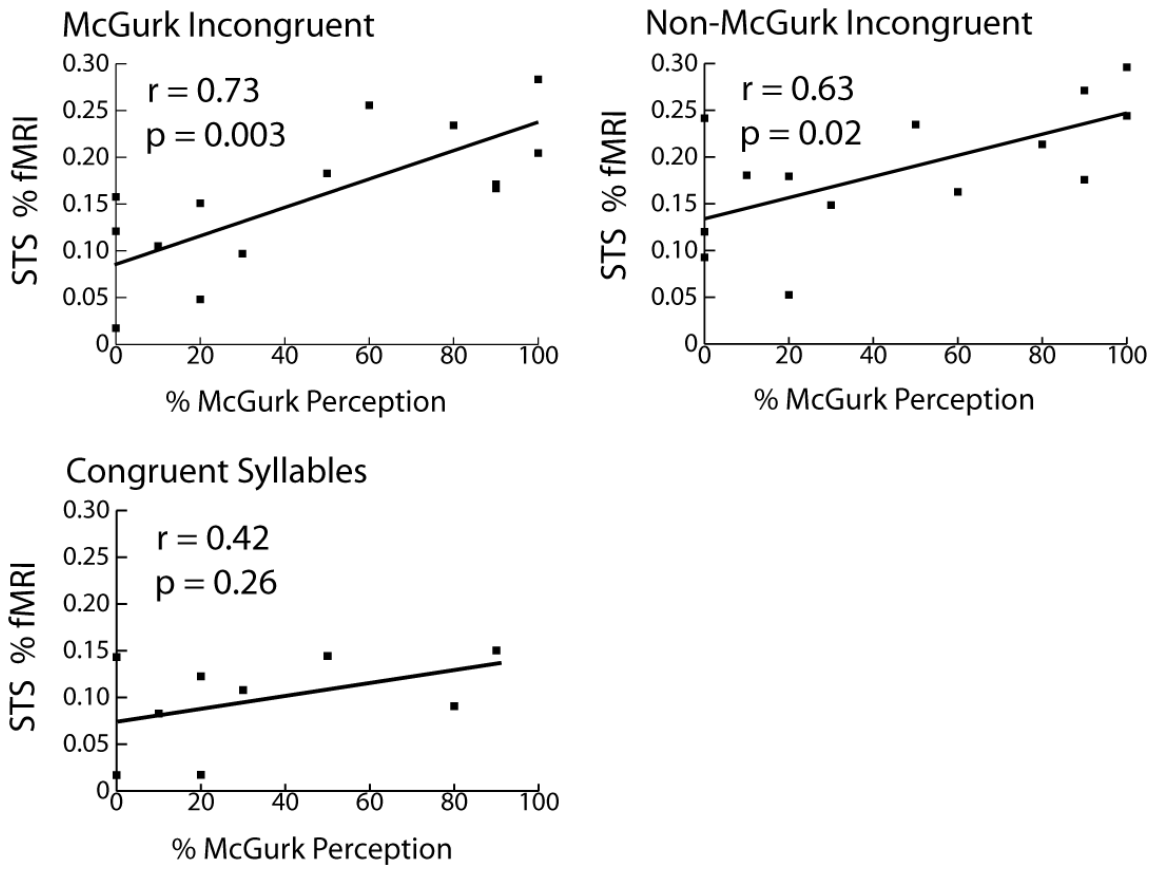

Figure 3.4 STS responses vs. McGurk susceptibility across subjects

The STS response to McGurk stimuli, non-McGurk incongruent stimuli and congruent stimuli in each subject are plotted against that subject's McGurk susceptibility.

Subjects with high and low STS responses to McGurk stimuli both identified target syllables with high precision (98% accuracy), indicating that both groups of subjects attended to the audiovisual stimuli. Furthermore, both groups showed similar STS responses to congruent syllables (0.12% vs. 0.08%, $p = 0.39$), indicated that there was not a systematic difference in attention or arousal between groups that modulated the STS response to all stimuli. Across subjects, there was no correlation between the STS response to congruent syllables and susceptibility ($r = 0.42$, $p = 0.26$).

Other Regions of interest

To examine other brain regions, we used our independent speech perception localizers to create three additional ROIs: Broca's area, auditory cortex and extrastriate visual cortex (Figure 3.5). Across these ROIs, there was no significant correlation between ROI activity and McGurk susceptibility for any stimulus condition (Broca's area: $r = -0.04$, $p = 0.89$ across incongruent stimuli; $r = 0.22$, $p = 0.57$ across congruent stimuli; auditory cortex: $r = 0.48$, $p = 0.08$; $r = 0.59$, $p = 0.10$; visual cortex: $r = -0.07$, $p = 0.81$; $r = -0.07$, $p = 0.86$).

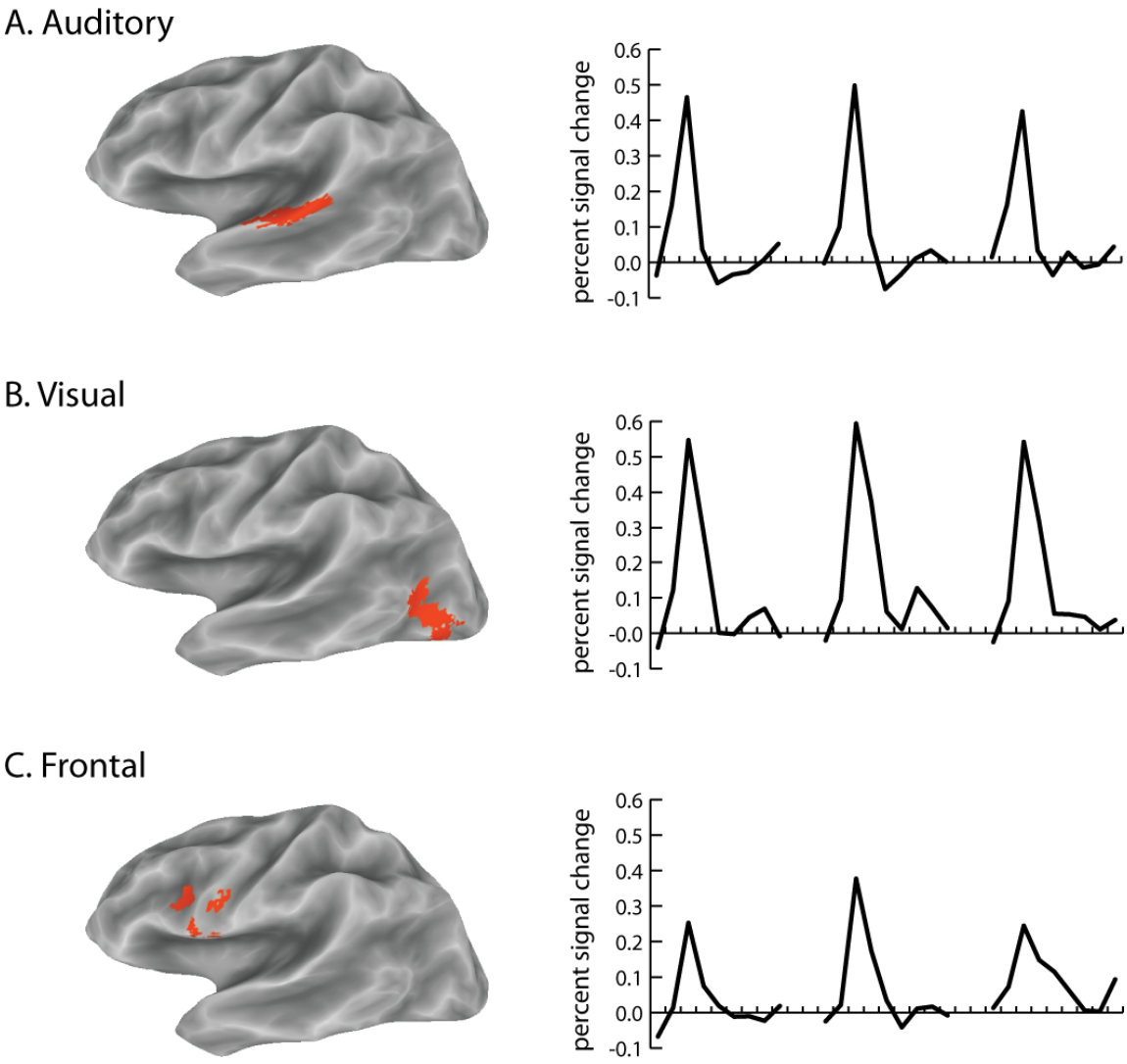

Figure 3.5 Cortical responses in other regions of interest

A. Hemodynamic responses to McGurk, non-McGurk and congruent stimuli (curves shown left to right) in the auditory cortex of one subject.
B. Hemodynamic responses to McGurk, non-McGurk and congruent stimuli in the visual cortex of the same subject.
C. Hemodynamic responses to McGurk, non-McGurk and congruent stimuli in the frontal cortex of the same subject.

Group Analyses

As an additional search for areas important for McGurk susceptibility, we performed a voxel-wise whole-brain group analysis. Results of this regression analysis showed no areas with a significant correlation with perception. Because this negative result contrasted sharply with the results of the ROI analysis, we investigated further. While each subject showed a large region of STS that responded to both auditory-only and visual-only speech stimuli, after transformation into standard space the overlap across subjects was very small, reflecting both anatomical and functional variability in the location of the STS multisensory area. To quantify this variability, we measured the location of the STS ROI in each subject. The mean (+- SD) center-of-mass was x = -53.8 +- 8.3 mm, y = -27.5 +- 9.6 mm, z = 3.5 +- 7.9 mm (Table 3.1). On average, each subject's STS center-of-mass was 13.7 mm from the mean center-of-mass, with some subjects more than 2 cm from the mean center-of-mass.

Average STS location				
		Talairach Coordinates		
ROI	# Voxels	x	y	z
STS	46.4 +- 28.5	-53.8 +- 8.3	-27.5 +- 7.9	3.5 +- 9.6

Table 3.1 Locations of STS across all subjects

We considered whether changes in functional connectivity between the STS and frontal cortex, auditory cortex or extrastriate visual cortex could predict behavioral perception of McGurk stimuli. No correlation was observed between McGurk susceptibility and STS-frontal cortex connectivity ($r = -0.31$, $p = 0.28$), STS-auditory cortex connectivity ($r = 0.41$, $p = 0.15$) or STS-visual cortex connectivity ($r = 0.34$, $p = 0.23$) during perception of McGurk stimuli.

Conclusions

To understand the neural basis for intersubject variability in the perception of the McGurk effect, we examined 14 subjects with a broad range of susceptibility to the McGurk effect (0% to 100%). Across subjects, we found a correlation between the amount of activity in the posterior STS during presentation of McGurk stimuli and subjects' susceptibility to the McGurk effect. The creation of a McGurk percept requires the integration of auditory and visual information: without the conflicting visual information, only the auditory syllable is perceived.

Many studies have identified the STS as a critical brain locus for auditory-visual integration for both speech and non-speech stimuli (20, 39-44, 148). An important role for the STS in the McGurk effect is supported by a recent TMS study, which demonstrated that interrupting activity in the STS significantly reduced the McGurk effect in those subjects who normally experience it (47). Disrupting the STS made these McGurk-susceptible individuals more similar to those of McGurk-resistant individuals: they were much more likely to perceive only the auditory-syllable of the McGurk stimulus.

This suggests a parsimonious explanation for the correlation between STS activity and McGurk susceptibility across individual. The posterior STS integrates auditory and visual information during speech perception. STS activity indicates that neural integration of auditory and visual information is occurring, resulting in the McGurk percept. If STS activity is reduced, auditory-visual integration does not occur and there is no McGurk percept.

Most studies of the McGurk effect report the mean probability of a McGurk percept across subjects and trials. Calculated in this way, in our dataset we found a McGurk probability of 46%, within the range reported in the literature: from 32% (Sams et al., 1998) to 49% (Benoit et al., 2010) to 64% (Bovo et al., 2009) to 79% (Baynes et al., 1994) to 83% (Olson et al., 2002) to 94% (Norrix et al., 2006). However, this grand mean probability conflates the intrasubject and intersubject variability: a grand mean probability of 50% could be explained by identical subjects, each of whom perceives the effect on half the trials; or by a distribution in which some always perceive the effect and some never do. Our results support the latter view. We found a dramatic range in the frequency of the McGurk percept across subjects from 0% to 100%. Only 36% of our subjects were highly susceptible to the illusion (>80% within-subject percept probability). While the initial report of the illusion claimed that 98% of subjects experienced the illusion (McGurk and MacDonald, 1976), recent studies have found much lower rates. Using the same threshold (>80% within-subject percept probability), Benoit et al. (2010) found a population likelihood of 31%. Two studies (thresholds not stated) reported population likelihoods of 26% (Gentilucci and Cattaneo, 2005) and 50% (MacDonald et al., 2000). Taken together, these results suggest that some subjects are highly susceptible to the McGurk effect and others are not.

What could explain this high degree of intersubject difference in McGurk susceptibility? One clue is found in the STS response to incongruent (non-McGurk) stimuli. Although these stimuli did not produce a fused percept in any individual, there was significant variation in the STS response to these incongruent stimuli across subjects. This variation in response to incongruent stimuli was significantly correlated

with McGurk susceptibility ($p = 0.02$). Multisensory integration uses independent sources of information from different sensory modalities to make more accurate judgments about the world. For multisensory integration to be beneficial, only information from the same stimuli should be integrated: a weak sound and a weak flash from the same location at the same time are independent evidence that an object is present, while a sound and a flash from different locations at different times are not; a similar argument holds for auditory and visual speech. The criteria used to determine whether auditory and visual speech should be integrated or not may be more or less stringent between individuals. Individuals with less stringent criteria (who could be thought of as possessing a more "forgiving" STS) attempt to integrate even obviously incongruent audiovisual speech. This produces activation in the STS for incongruent stimuli, and the McGurk percept for McGurk syllables. Individuals with more stringent criteria (less forgiving STS) to not attempt to integrate incongruent audiovisual speech and do not perceive the McGurk effect. An obvious and important question for future research is to determine if the stringency of criteria for multisensory integration extends to other stimulus manipulations, such as differences in the timing of auditory and visual speech, or in the noise present in the auditory or visual modalities. If subjects could be trained to change the stringency of their criteria for multisensory integration, for instance using neurofeedback (158, 159), then behavioral measures of multisensory integration, such as McGurk susceptibility, might show a concomitant increase. This could be useful for treating patients with language deficits such as dyslexia (160, 161) or patients with cochlear implants who do not integrate auditory speech with visual lip movements as strongly as people with normal hearing (162). The stringency of criteria for multisensory

integration may change with development. Children are less susceptible to the McGurk effect (11, 163). We would predict that STS activity would be diminished in children, accounting for their decreased audiovisual integration.

Our studies add to a growing body of literature relating differences in brain function to differences in individual language abilities (21, 33-35, 164). For instance, Hall et al. (2005) studied individual differences in visual speech-reading and found that subjects with greater speech-reading performance had a greater number of activated voxels in the left superior temporal gyrus during an auditory comprehension task. Wong et al. (2007) found that areas of the left posterior STS showed increased activation in subjects who more readily acquired tone patterns in a novel tone-based language, while right-sided areas including the right posterior STS showed increased activation in the subjects who had more difficulty in learning these pitch patterns. Mei et al. (2008) studied native Chinese speakers who were trained to learn an artificial language and found increased activity in left middle temporal gyrus and STS for the participants who showed above-average behavioral performance than those who were below average. Aziz-Zadeh et al. (2010) studied areas that responded processing of highly prosodic speech. One of the areas that was responsive during perception of prosodic speech, the left IFG, showed greater activity in subjects with higher behavioral scores in an empathy task. This finding indicates that heightened activity of a prosodic area may subserve the ability to use social cues involved in detecting distress of others. Eisner et al. (2010) found that subjects who were better able to learn to recognize noise-vocoded words after training exhibited greater activity in the left IFG during these noisy auditory stimuli. Taken together with these results, our findings support the notion that increased cortical

activity in language-related areas may be predictive of inter-subject differences in speech perception.

Finally, we turn to the question of why the strong correlation that we found between McGurk susceptibility and STS response has not been observed previously. There have a number of previous studies of the McGurk effect using lesion data (165, 166), EEG (167-170), MEG (171-173), PET (20) and fMRI (149-152, 174, 175). Four studies differentiated subjects based on their susceptibility to the McGurk effect (149, 150, 152, 175) but failed to find the positive correlation between STS activity and McGurk susceptibility observed in our experiment. A possible explanation for this failure is that none of the previous studies used independent functional localizers to identify the STS.

The study by Jones and Callan (2003) used voxel-wise regression on fMRI data to search for voxels with a significant correlation between brain activity and McGurk susceptibility. No correlation in any STS voxels was reported. Similarly, in another study by Wiersinga-Post et al. (2010), voxel-wise regression was used to identify cortical regions that showed significant correlation between BOLD signal and degree of McGurk perception at five different audiovisual delays. As with the Jones and Callan result, there were no areas which showed a positive correlation between activity and McGurk perception. However, because of intersubject variability, examining individual voxels in standard space may not compare functionally homologous regions between subjects. In the present study, the STS multisensory area was more than 2 cm from the mean location in some subjects (43, 72, 176). Because of this high intersubject variability (and consistent with the findings of Jones and Callan and Wiersinga-Post et

al.), a voxel-wise ANOVA on our data did not reveal a correlation between STS activity and McGurk susceptibility.

A study by Hasson et al. (2007) used a repetition suppression paradigm to examine the fMRI response to different congruent syllables followed by a McGurk syllable. No differences between conditions were reported in the STS. However, the primary regions of interest (ROI) were defined anatomically. This presents a problem when studying the STS because the STS is the second largest sulcus in the human brain, after the Sylvian fissure (177). Because the STS multisensory area constitutes only a small portion of the entire STS, averaging across all voxels in the STS includes many voxels that have no response to speech stimuli, decreasing statistical power. This effect is illustrated by the fMRI study of Benoit et al. (2010) that also used repetition suppression of McGurk stimuli with anatomical ROIs. Our reanalysis of the Benoit et al. data (Figure 3) found that the STS response to McGurk stimuli (using an anatomical ROI consisting of the entire STS) was not significantly different from zero ($p = 0.90$). In contrast, in our data, the STS response to McGurk stimuli (using an ROI from the independent functional localizer) was significantly greater than zero (mean response of 0.16%, $p = 0.000003$). Benoit et al. (2010) reported an *inverse* relationship between McGurk susceptibility and activity in the STS, the exact opposite of our effect. While the mean response of the STS in Benoit et al. study was not significantly different from zero, individual subjects had very large signal changes, with signal changes of -1%, -2% and -4% in the STS of the three subjects with the highest McGurk susceptibility. These signal changes are both an order of magnitude larger than in previous studies (e.g. mean response of 0.16% in the present study) and in the wrong direction: previous studies in

the literature report *positive* responses to audiovisual speech in the STS (20, 40, 44, 106, 178). Therefore, it seems likely that these large negative signal changes are an artifact of the anatomically-defined STS ROI used by Benoit et al.

In summary, previous studies did not use functional localizers to identify the location of the multisensory portion of STS in each individual subject. Using functional localizers, we found a strong relationship between STS activity and McGurk susceptibility.

CHAPTER 4: CONCLUSIONS AND FUTURE DIRECTIONS

Audiovisual integration is a critical component of understanding speech, but the brain mechanisms that underlie this process are not completely understood. By clarifying the role of an important multisensory cortical region, the multisensory posterior STS, we can better understand how the brain integrates auditory and visual components of speech during speech comprehension. The purpose of the first set of experiments was to understand the mechanism by which the STS integrates information from connected auditory and visual areas. We were interested in the interactions between STS and auditory and visual areas, and if the STS integrates these inputs in a weighted manner depending on the quality of information in each input stream. For example, in a noisy room, we will use more information from the more reliable visual modality than the less reliable auditory modality, and our audiovisual perception in that setting is more dependent on the visual input. Is this perceptual reliability-weighting a product of a reliability-weighting process carried out by the STS? We found that the neurons within the STS weight the auditory and visual inputs they receive, and the STS correlates its activity with the more reliable modality.

We next aimed to understand the role of STS activity in audiovisual perception: does this activity within the STS predict how strongly a person integrates auditory and visual speech information? In order to clarify how brain activity within an individual's STS correlates with that person's audiovisual perception, we studied subjects' perception of audiovisual McGurk syllables as well as the amplitude of cortical response within the STS as measured by fMRI during the audiovisual McGurk illusion as well as during syllables not associated with any audiovisual illusion. We found that subjects who perceive the McGurk illusion more strongly have a correlated increase in amplitude of

response of the multisensory STS. Taken together, these results provide evidence that activity within the left posterior STS is critical in the integration of auditory and visual components of speech.

Next we consider the possible clinical relevance of these basic research findings. First, we examine the relevance for stroke patients. Next, we examine the relevance for healthy aging. Finally, we consider the relevance for developmental language disorders.

These studies may give us insight into the progression of recovery after brain damage due to cerebrovascular infarcts and excision of tumors or foci of epileptiform activity. For example, Hamilton et al. (2006) describe the case of a patient who underwent a diffuse stroke, including temporal and parietal areas, who subsequently lost the ability to integrate mouth movements with auditory speech. As a result, he would turn away from a speaker's face during conversation, and he preferred communication by telephone in order to avoid the now distracting visual speech stream. In addition to his difficulties with everyday conversation, he also did not perceive the McGurk effect. For patients with a similar loss of multisensory integration, it may be useful to monitor their recovery using serial tests of McGurk perception. In conjunction with these behavioral methods, multiple measures of STS activity using fMRI and quantification of increases in connectivity with auditory and visual cortical areas may provide evidence of recovery of pathways for integrating auditory and visual information.

A large proportion of the aging population will face individual sensory losses from age-related hearing and vision loss (179, 180). As hearing declines, visual input from mouth movements must be used more efficiently to compensate for the auditory deficit, and vice versa. Over time, these changes in the external reliability of the sensory

input may be reflected in the cortex as a decreased connection weight between the STS and the sensory cortex processing the noisier input.

Characterizing the activity of multisensory STS and its interaction with connected early sensory areas opens the door to studying the manner in which audiovisual integration changes in normal development. From the behavioral literature, there is evidence that the ability to integrate visual mouth movements with auditory speech in children strengthens with age (11, 163, 181-183). In a study of children ranging in age from 5 to 12 years by Tremblay et al. (2007), it was found that perception of the audiovisual McGurk effect was greater for the children aged 10-12 years than the children aged 5-9 years. It would be fascinating to determine if this increase in audiovisual integration in development is subserved by an increase in STS activity or a change in connectivity over time.

In addition to studying the development of audiovisual integration in the setting of typical development, the quantification of STS activity and connectivity may allow us to monitor changes in multisensory brain activity during rehabilitation from a number of sensory disorders of development. In the context of neurological disorders, such as children with hearing impairment undergoing cochlear implantation, it would be advantageous to be able to monitor how and when the brain begins to integrate this newly-perceived auditory speech with the familiar visual speech. By characterizing changes in functional connectivity between STS and auditory areas in normal subjects during audiovisual speech with different auditory noise levels, we may build a model for understanding how audiovisual integration changes with the addition of more reliable auditory input.

Additionally, characterizing changes in multisensory activity in the STS may help us better understand and treat a common disorder of reading, dyslexia. Dyslexia is one of the most common learning disorders in the United States. It is estimated that as many as 5% of children may have dyslexia (184). In both affected children and adults, reading performance is poor despite normal intelligence, motivation and schooling. In addition to having problems with reading, it has been found that there are differences in integrating information from different sensory modalities in this population. Hairston et al. (2005) studied auditory-visual multisensory integration of auditory noise bursts and visual circles in dyslexic and typical readers (160). They found that dyslexic readers integrated auditory and visual stimuli over longer time periods than the controls, showing evidence of faulty temporal binding of auditory and visual cues in dyslexics. This deficit in multisensory integration may underlie difficulties in reading; if a visual word cannot be paired with the matching auditory pronunciation in a reasonable time window, then reading is impaired. In a study of dyslexic and typical readers by Pekkola et al. (2006), dyslexic readers were found to have more extensive activation during conflicting audiovisual speech (such as auditory /o/ with visual /i/) in motor speech regions speech regions such as left inferior parietal lobule and supplementary motor area (185). This finding suggests processing of multisensory speech requires additional motor loop processing in dyslexics to overcome deficits in multisensory processing.

The only diagnostic method currently available for dyslexia is behavioral testing, which is problematic due to the time-consuming nature of neuropsychological testing. When the diagnosis of dyslexia is delayed, affected adolescents and young adults are more likely to drop out of school and face legal and psychiatric problems. Given these

comorbidities, it would be advantageous to develop a rapid method of diagnosis that could also be used to monitor progress during remediation. By studying STS activity and connectivity with early sensory areas, we may be able to use fMRI, or other neuroimaging modalities such as near-infrared spectroscopy (186), in the future as a rapid diagnostic imaging tool in assessing the extent of multisensory deficits in dyslexics and quantifying improvement in multisensory processing after therapy. In order to characterize abnormalities in white matter connectivity in dyslexic subjects (187), structural equation modeling could be used to identify the direction and strength of neural pathways during language processing. It could be hypothesized that unlike normal controls, dyslexic readers will not have heightened connection strengths during audiovisual language processing. Once patterns of fMRI activation and connectivity associated with multisensory integration in dyslexics have been identified and distinguished from controls, perhaps fMRI may be used in the future as a rapid diagnostic imaging tool in assessing the extent of multisensory deficits in dyslexics and quantifying improvement in multisensory processing after therapy.

BIBLIOGRAPHY

1. Sumby, W. H., and I. Pollack. 1954. Visual contribution to speech intelligibility in noise. Journal of the Acoustic Society of America 26:212-215.
2. MacLeod, A. M., and Q. Summerfield. 1990. A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: rationale, evaluation, and recommendations for use. British Journal of Audiology 24:29-43.
3. Ross, A. R., D. Saint-Amour, V. M. Leavitt, D. C. Javitt, and J. J. Foxe. 2007. Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. Cereb Cortex 17:1147-1153.
4. MacLeod, A. M., and A. Q. Summerfield. 1990. A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: rationale, evaluation, and recommendations for use. British Journal of Audiology 24:29-43.
5. Risberg, A., and J. L. Lubker. 1978. Prosody and speechreading. Speech Transmission Laboratory Quarterly Progress & Status Report 4:1-16.
6. Remez, R. E., J. M. Fellowes, D. B. Pisoni, W. D. Goh, and P. E. Rubin. 1998. Multimodal perceptual organization of speech: Evidence from tone analogs of spoken utterances. Speech Communication 16:65-73.
7. Alais, D., and D. Burr. 2004. The ventriloquist effect results from near-optimal bimodal integration. Curr Biol 14:257-262.
8. Ernst, M. O., and M. S. Banks. 2002. Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415:429-433.

9. Witten, I. B., and E. I. Knudsen. 2005. Why seeing is believing: merging auditory and visual worlds. Neuron 48:489-496.

10. Ma, W. J., X. Zhou, L. A. Ross, J. J. Foxe, and L. C. Parra. 2009. Lip-reading aids word recognition most in moderate noise: A Bayesian explanation using high-dimensional feature space. PLoS ONE 4.

11. McGurk, H., and J. W. MacDonald. 1976. Hearing lips and seeing voices. Nature 264:746-748.

12. Hackett, T. A., T. M. Preuss, and J. H. Kaas. 2001. Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. J Comp Neurol 441:197-222.

13. Liegeois-Chauvel, C., J. B. de Graaf, V. Laguitton, and P. Chauvel. 1999. Specialization of left auditory cortex for speech perception in man depends on temporal coding. Cereb Cortex 9:484-496.

14. Scott, S. K., and I. S. Johnsrude. 2003. The neuroanatomical and functional organization of speech perception. Trends Neurosci 26:100-107.

15. Belin, P., R. J. Zatorre, and P. Ahad. 2002. Human temporal-lobe response to vocal sounds. Brain Res Cogn Brain Res 13:17-26.

16. Okada, K., F. Rong, J. Venezia, W. Matchin, I.-H. Hsieh, K. Saberi, J. T. Serences, and G. Hickok. 2010. Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. Cereb Cortex 20:2486-2495.

17. Poeppel, D., C. Wharton, J. Fritz, A. Guillemin, L. San Jose, J. Thompson, D. Bavelier, and A. Braun. 2004. FM sweeps, syllables and word stimli differentially modulate left and right non-primary auditory areas. Neuropsychologia 42:183-200.

18. Hickok, G. 2009. The functional neuroanatomy of language. Phys Life Rev 6:121-143.

19. Ludman, C. N., A. Q. Summerfield, D. Hall, M. R. Elliott, J. Foster, J. L. Hykin, R. Bowtell, and P. G. Morris. 2000. Lip-reading ability and patterns of cortical activation studied using fMRI. Br J Audiol 34:225-230.

20. Sekiyama, K., I. Kanno, S. Miura, and Y. Sugita. 2003. Auditory-visual speech perception examined by fMRI and PET. Neurosci Res 47:277-287.

21. Hall, D. A., C. Fussell, and A. Q. Summerfield. 2005. Reading fluent speech from talking faces: typical brain networks and individual differences. J Cogn Neurosci 17:939-953.

22. Puce, A., T. Allison, S. Bentin, J. C. Gore, and G. McCarthy. 1998. Temporal cortex activation in humans viewing eye and mouth movements. J Neurosci 18:2188-2199.

23. Ruytjens, L., F. Albers, P. van Dijk, and A. Willemsen. 2006. Neural responses to silent lipreading in normal hearing male and female subjects. Eur J Neurosci 24:1835-1844.

24. Nishitani, N., and R. Hari. 2002. Viewing lip forms: cortical dynamics. Neuron 36:1211-1220.

25. Keller, S. S., T. Crow, A. Foundas, K. Amunts, and N. Roberts. 2009. Broca's area: nomenclature, anatomy, typology and asymmetry. Brain Lang 109:29-48.

26. Ojanen, V., R. Mottonen, J. Pekkola, I. P. Jaaskelainen, R. Joensuu, T. Autti, and M. Sams. 2005. Processing of audiovisual speech in Broca's area. NeuroImage 25:333-338.

27. Broca, P. 2006. Comments regarding the seat of the faculty of spoken language, followed by an observation of aphemia (loss of speech). Oxford University Press, New York.

28. Schnur, T. T., M. F. Schwartz, D. Y. Kimberg, E. Hirshorn, H. B. Coslett, and S. L. Thompson-Schill. 2009. Localizing interference during naming: convergent neuroimaging and neuropsychological evidence for the function of Broca's area. Proc Natl Acad Sci U S A 106:322-327.

29. Wise, R. J. S., S. K. Scott, S. C. Blank, C. J. Mummery, K. Murphy, and E. A. Warburton. 2001. Separate neural subsystems within 'Wernicke's area'. Brain 124:83-95.

30. Wernicke, C. 1968. The symptom complex of aphasia (1874). Proc. Boston Colloq. Philos. Sci. 4:34-97.

31. Poeppel, D., W. J. Idsardi, and V. van Wassenhove. 2008. Speech perception at the interface of neurobiology and linguistics. Philosophical Transactions of the Royal Society B 363:1071-1086.

32. Hickok, G., and D. Poeppel. 2007. The cortical organization of speech perception. Nature Reviews Neuroscience 8:393-402.

33. Wong, P. C., T. K. Perrachione, and T. B. Parrish. 2007. Neural characteristics of successful and less successful speech and word learning in adults. Hum Brain Mapp 28:995-1006.

34. Mei, L., C. Chen, G. Xue, Q. He, T. Li, F. Xue, Q. Yang, and Q. Dong. 2008. Neural predictors of auditory word learning. Neuroreport 19:215-219.

35. Eisner, F., C. McGettigan, A. Faulkner, S. Rosen, and S. K. Scott. 2010. Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. J Neurosci 30:7179-7186.

36. Kayser, C., and N. Logothetis. 2009. Directed interactions between auditory and superior temporal cortices and their role in sensory integration. Frontiers in Integrative Neuroscience 3:1-11.

37. Chandrasekaran, C., and A. A. Ghazanfar. 2009. Different neural frequency bands integrate faces and voices differently in the superior temporal sulcus. J Neurophysiol 101:773-788.

38. Reale, R. A., G. A. Calvert, T. Thesen, R. L. Jenison, H. Kawasaki, H. Oya, M. A. Howard, and J. F. Brugge. 2007. Auditory-visual processing represented in the human superior temporal gyrus. Neuroscience 145:162-184.

39. Callan, D. E., J. A. Jones, K. Munhall, C. Kroos, A. M. Callan, and E. Vatikiotis-Bateson. 2004. Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. J Cogn Neurosci 16:805-816.

40. Stevenson, R. A., and T. W. James. 2009. Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. Neuroimage 44:1210-1223.

41. Calvert, G. A., R. Campbell, and M. J. Brammer. 2000. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr Biol 10:649-657.

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

43. Beauchamp, M. S., K. E. Lee, B. D. Argall, and A. Martin. 2004. Integration of auditory and visual information about objects in superior temporal sulcus. Neuron 41:809-823.

44. Wright, T. M., K. A. Pelphrey, T. Allison, M. J. McKeown, and G. McCarthy. 2003. Polysensory interactions along lateral temporal regions evoked by audiovisual speech. Cereb Cortex 13:1034-1043.

45. Seltzer, B., M. G. Cola, C. Gutierrez, M. Massee, C. Weldon, and C. G. Cusick. 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.

46. Lewis, J. W., and D. C. Van Essen. 2000. Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. J Comp Neurol 428:112-137.

47. Beauchamp, M. S., A. R. Nath, and S. Pasalar. 2010. fMRI-guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. J Neurosci 30:2414-2417.

48. Stein, B. E., and M. A. Meredith. 1993. The Merging of the Senses. MIT Press.

49. Ma, W. J., J. M. Beck, P. E. Latham, and A. Pouget. 2006. Bayesian inference with probabilistic population codes. Nat Neurosci 9:1432-1438.

50. Werner, S., and U. Noppeney. 2009. Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. Cereb Cortex.

51. Dahl, C. D., N. K. Logothetis, and C. Kayser. 2009. Spatial organization of multisensory responses in temporal association cortex. J Neurosci 29:11924-11932.

52. Beauchamp, M. S. 2005. See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. Curr Opin Neurobiol 15:145-153.

53. Van Essen, D. C. 2005. A Population-Average, Landmark- and Surface-based (PALS) atlas of human cerebral cortex. Neuroimage 28:635-662.

54. Hein, G., and R. T. Knight. 2008. Superior temporal sulcus--It's my area: or is it? J Cogn Neurosci 20:2125-2136.

55. Campbell, R. 2008. The processing of audio-visual speech: empirical and neural bases. Phil. Trans. R. Soc. B 363:1001-1010.

56. Price, C. J. 2000. The anatomy of language: contributions from functional neuroimaging. J Anat 197 Pt 3:335-359.

57. Binder, J. R., J. A. Frost, T. A. Hammeke, R. W. Cox, S. M. Rao, and T. Prieto. 1997. Human brain language areas identified by functional magnetic resonance imaging. J Neurosci 17:353-362.

58. Belin, P., S. Fecteau, and C. Bedard. 2004. Thinking the voice: neural correlates of voice perception Trends in Cognitive Sciences 8:129-135.

59. Zatorre, R. J. 2007. There's more to auditory cortex than meets the ear. Hear Res 229:24-30.

60. McIntosh, A. R., and F. Gonzalez-Lima. 1994. Structural equation modeling and its application to network analysis in functional brain imaging. Human Brain Mapping 2:2-22.

61. Buchel, C., and K. Friston. 2001. Interactions among neuronal systems assessed with functional neuroimaging. Rev Neurol (Paris) 157:807-815.

62. Horwitz, B. 2003. The elusive concept of brain connectivity. NeuroImage 19:466-470.

63. Stein, J. L., L. M. Wiedholz, D. S. Bassett, D. R. Weinberger, C. F. Zink, V. S. Mattay, and A. Meyer-Lindenberg. 2007. A validated network of effective amygdala connectivity. Neuroimage 36:736-745.

64. de Marco, G., P. Vrignaud, C. Destrieux, D. de Marco, S. Testelin, B. Devauchelle, and P. Berquin. 2009. Principle of structural equation modeling for exploring functional interactivity within a putative network of interconnected brain areas. Magn Reson Imaging 27:1-12.

65. Wilson, M. 1988. The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2. Behavioral Research Methods, Instruments and Computers 20:6-11.

66. Shannon, R. V., F.-G. Zeng, V. Kamath, J. Syngonski, and M. Ekelid. 1995. Speech recognition with primarily temporal cues. Science 270:303-304.

67. Munhall, K., G., C. Kroos, G. Jozan, and E. Vatikiotis-Bateson. 2004. Spatial frequency requirements for audiovisual speech perception. Perception & Psychophysics 66:574-583.

68. Talairach, J., and P. Tournoux. 1988. Co-Planar stereotaxic atlas of the human brain. Thieme Medical Publishers, New York.

69. Mazziotta, J., A. Toga, A. Evans, P. Fox, J. Lancaster, K. Zilles, R. Woods, T. Paus, G. Simpson, B. Pike, C. Holmes, L. Collins, P. Thompson, D. MacDonald, M. Iacoboni, T. Schormann, K. Amunts, N. Palomero-Gallagher, S. Geyer, L. Parsons, K. Narr, N. Kabani, G. Le Goualher, D. Boomsma, T. Cannon, R. Kawashima, and B. Mazoyer. 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.

70. Fischl, B., M. I. Sereno, and A. M. Dale. 1999. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. Neuroimage 9:195-207.

71. Dale, A. M., B. Fischl, and M. I. Sereno. 1999. Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 9:179-194.

72. Argall, B. D., Z. S. Saad, and M. S. Beauchamp. 2006. Simplified intersubject averaging on the cortical surface using SUMA. Hum Brain Mapp 27:14-27.

73. Cox, R. W. 1996. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 29:162-173.

74. Genovese, C. R., N. A. Lazar, and T. Nichols. 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage 15:870-878.

75. Saad, Z. S., D. R. Glen, G. Chen, M. S. Beauchamp, R. Desai, and R. W. Cox. 2009. A new method for improving functional-to-structural MRI alignment using local Pearson correlation. Neuroimage 44:839-848.

76. Cohen, M. S. 1997. Parametric analysis of fMRI data using linear systems methods. Neuroimage 6:93-103.

77. Kriegeskorte, N., W. K. Simmon, P. S. Bellgowan, and C. I. Baker. 2009. Circular analysis in systems neuroscience: the dangers of double dipping. Nature Neuroscience 12:535-540.

78. Branch, D., B. Milner, and T. Rasmussen. 1964. Intracarotid sodium amytal for the lateralization of cerebral speech dominance; observations in 123 patients. Journal of Neurosurgery 21:399-405.

79. Ellmore, T. M., M. S. Beauchamp, J. I. Breier, J. D. Slater, G. P. Kalamangalam, T. J. O'Neill, M. A. Disano, and N. Tandon. 2010. Temporal lobe white matter asymmetry and language laterality in epilepsy patients. Neuroimage 49:2033-2044.

80. Upadhyay, J., T. A. Knaus, K. A. Lindgren, M. Ducros, D.-S. Kim, and H. Tager-Flusberg. 2008. Effective and structural connectivity in the human auditory cortex. J Neurosci 28:3341-3349.

81. Patterson, R. D., and I. S. Johnsrude. 2008. Functional imaging of the auditory processing applied to speech sounds. Philosophical Transactions of the Royal Society B 363:1023-1035.

82. Tootell, R. B., J. B. Reppas, K. K. Kwong, R. Malach, R. T. Born, T. J. Brady, B. R. Rosen, and J. W. Belliveau. 1995. Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. J Neurosci 15:3215-3230.

83. Beauchamp, M. S., R. W. Cox, and E. A. DeYoe. 1997. Graded effects of spatial and featural attention on human area MT and associated motion processing areas. J Neurophysiol 77:516-520.

84. Downing, P. E., Y. Jiang, M. Shuman, and N. Kanwisher. 2001. A cortical area selective for visual processing of the human body. Science 293:2470-2473.

85. Beauchamp, M. S., K. E. Lee, J. V. Haxby, and A. Martin. 2002. Parallel visual motion processing streams for manipulable objects and human movements. Neuron 34:149-159.

86. Beauchamp, M. S., K. E. Lee, J. V. Haxby, and A. Martin. 2003. fMRI Responses to Video and Point-Light Displays of Moving Humans and Manipulable Objects. J Cognit Neurosci 15:991-1001.

87. Pelphrey, K. A., J. P. Morris, C. R. Michelich, T. Allison, and G. McCarthy. 2005. Functional anatomy of biological motion perception in posterior temporal cortex: an fMRI study of eye, mouth and hand movements. Cereb Cortex.

88. Dumoulin, S. O., R. G. Bittar, N. J. Kabani, C. L. Baker, Jr., G. Le Goualher, G. Bruce Pike, and A. C. Evans. 2000. A new anatomical landmark for reliable identification of human area V5/MT: a quantitative analysis of sulcal patterning. Cereb Cortex 10:454-463.

89. Beauchamp, M. S., N. E. Yasar, R. E. Frye, and T. Ro. 2008. Touch, sound and vision in human superior temporal sulcus. Neuroimage 41:1011-1020.

90. Beauchamp, M. S. 2005. Statistical criteria in FMRI studies of multisensory integration. Neuroinformatics 3:93-114.

91. Ihaka, R., and R. Gentleman. 1996. R: A language for data analysis and graphics. Journal of Computational and Graphical Statistics 5:299-314.

92. Buchel, C., and K. J. Friston. 1997. Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. Cereb Cortex 7:768-778.

93. Chen, G., D. R. Glen, J. L. Stein, A. S. Meyer-Lindenberg, Z. S. Saad, and R. W. Cox. 2007. Model validation and automated search in fMRI path analysis: a fast open-source tool for structural equation modeling. In Human Brain Mapping Conference.

94. Friston, K. J., C. Buchel, G. R. Fink, J. Morris, E. T. Rolls, and R. J. Dolan. 1997. Psychophysiological and modulatory interactions in neuroimaging. Neuroimage 6:218-229.

95. Dale, A. M. 1999. Optimal experimental design for event-related fMRI. Hum Brain Mapp 8:109-114.

96. Friston, K. J., and C. Buchel. 2000. Attentional modulation of effective connectivity from V2 to V5/MT in humans. Proc Natl Acad Sci U S A 97:7591-7596.

97. Davis, C., and J. Kim. 2004. Audio-visual interactions with intact clearly audible speech. Q J Exp Psychol A 57:1103-1121.

98. Grant, K. W., and P. F. Seitz. 2000. The use of visible speech cues for improving auditory detection of spoken sentences. J Acoust Soc Am 108:1197-1208.

99. Shahin, A. J., and L. M. Miller. 2009. Multisensory integration enhances phonemic restoration. J Acoust Soc Am 125:1744-1750.

100. Laurienti, P. J., R. A. Kraft, J. A. Maldjian, J. H. Burdette, and M. T. Wallace. 2004. Semantic congruence is a critical factor in multisensory behavioral performance. Exp Brain Res 158:405-414.

101. MacDonald, J. D., S. Andersen, and T. Bachmann. 2000. Hearing by eye: how much spatial degradation can be tolerated? Perception 29:1155-1168.

102. Bitterman, Y., R. Mukamel, R. Malach, I. Fried, and I. Nelken. 2007. Ultra-fine frequency tuning revealed in single neurons of human auditory cortex. Nature 451:197-201.

103. Phillips, D. P., and D. R. F. Irvine. 1981. Responses of single neurons in physiologically defined primary auditory cortex (AI) of the cat: frequency tuning and responses to intensity. J Neurophysiol 45:48-58.

104. Recanzone, G. H. 2000. Response profiles of auditory cortical neurons to tones and noise in behaving macaque monkeys. Hearing Research 150:104-118.

105. Davis, M. H., and I. S. Johnsrude. 2003. Hierarchical processing in spoken language comprehension. J Neurosci 23:3423-3431.

106. Giraud, A. L., C. A. Kell, C. Thierfelder, P. Sterzer, M. O. Russ, C. Preibisch, and A. Kleinschmidt. 2004. Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. Cereb Cortex 14:247-255.

107. Hubel, D. H., and T. N. Wiesel. 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol 160:106-154.

108. Macknik, S. L., S. Martinez-Conde, and M. M. Haglund. 2000. The role of spatiotemporal edges in visibility and visual masking. Proc Natl Acad Sci U S A 97:7556-7560.

109. Albrecht, D. G., and D. B. Hamilton. 1982. Striate cortex of monkey and cat: contrast response function. J Neurophysiol 48:217-237.

110. Contreras, D., and L. Palmer. 2003. Response to contrast of electrophysiologically definted cell classes in primary visual cortex. J Neurosci 23:6936-6945.

111. Olman, C. A., K. Ugurbil, P. Schrater, and D. Kersten. 2004. BOLD fMRI and psychophysical measurements of contrast response to broadband images. Vision Res 44:669-683.

112. Park, J. C., X. Zhang, J. Ferrera, J. Hirsch, and D. C. Hood. 2008. Comparison of contrast-response functions from multifocal visual-evoked potentials (mfVEPs) and functional MRI responses. Journal of Vision 8:1-12.

113. Beauchamp, M. S., A. R. Nath, and S. Pasalar. 2010. fMRI-Guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. J Neurosci 30:2414-2417.

114. Scott, S. K., and I. S. Johnsrude. 2003. The neuroanatomical and functional organization of speech perception. Trends Neurosci 26:100-107.

115. Beauchamp, M. S., S. Pasalar, and T. Ro. 2010. Neural substrates of reliability-weighted visual-tactile multisensory integration. Frontiers in Systems Neuroscience 4.

116. Van Atteveldt, N., A. Roebroeck, and R. Goebel. 2009. Interaction of speech and script in human auditory cortex: insights from neuro-imaging and effective connectivity. Hearing Research 258:152-164.

117. Murray, S. O., D. Kersten, B. A. Olshausen, P. Schrater, and D. L. Woods. 2002. Shape perception reduces activity in human primary visual cortex. Proc Natl Acad Sci U S A 99:15164-15169.

118. Felleman, D. J., and D. C. Van Essen. 1991. Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1:1-47.

119. de la Mothe, L. A., S. Blumell, Y. Kajikawa, and T. A. Hackett. 2006. Cortical connections of the auditory cortex in marmoset monkeys: Core and medial belt regions. J Comp Neurol 496:27-71.

120. Winer, J. A. 2006. Decoding the auditory corticofugal systems. Hearing Research 212:1-8.

121. Smiley, J. F., T. A. Hackett, I. Ulbert, G. Karmas, P. Lakatos, D. C. Javitt, and C. E. Schroeder. 2007. Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. J Comp Neurol 502:894-923.

122. Seltzer, B., and D. N. Pandya. 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.

123. Kanwisher, N., and G. Yovel. 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.

124. von Kriegstein, K., O. Dogan, M. Gruter, A. L. Giraud, C. A. Kell, T. Gruter, A. Kleinschmidt, and S. J. Keibel. 2008. Simulation of talking faces in the human brain improves auditory speech recognition. Proceedings of the National Academy of Sciences 105:6747-6752.

125. Dan, Y., and M. Poo. 2004. Spike timing-dependent plasticity of neural circuits. Neuron 44:23-30.

126. Lee, J. Y., and J. H. Maunsell. 2009. A normalization model of attentional modulation of single unit responses. PLoS ONE 4:e4651.

127. Reynolds, J. H., and D. J. Heeger. 2009. The normalization model of attention. Neuron 61:168-185.

128. Kersten, D., P. Mamassian, and A. Yuille. 2004. Object perception as Bayesian inference. Annu. Rev. Psychol. 55:271-304.

129. Hesselmann, G., S. Sadaghiani, K. J. Friston, and A. Kleinschmidt. 2010. Predictive coding or evidence accumulation? False inference and neuronal fluctuations. PLoS ONE 5:e9926.

130. Ma, W. J., and A. Pouget. 2008. Linking neurons to behavior in multisensory perception: a computational review. Brain Res 1242:4-12.

131. Morgan, M. L., G. C. Deangelis, and D. E. Angelaki. 2008. Multisensory integration in macaque visual cortex depends on cue reliability. Neuron 59:662-673.

132. Fetsch, C. R., G. C. Deangelis, and D. E. Angelaki. Visual-vestibular cue integration for heading perception: applications of optimal cue integration theory. Eur J Neurosci 31:1721-1729.

133. Senkowski, D., T. R. Schneider, J. J. Foxe, and A. K. Engel. 2008. Crossmodal binding through neural coherence: implications for multisensory processing. Trends Neurosci 31:401-409.

134. Engel, A. K., P. Fries, and W. Singer. 2001. Dynamic predictions: oscillations and synchrony in top-down processing. Nature Reviews Neuroscience 2:704-716.

135. Chawla, D., E. D. Lumer, and K. Friston. 2000. Relating macroscopic measures of brain activity to fast, dynamic neuronal interactions. Neural Computation 12:2805-2821.

136. Helbig, H. B., and M. O. Ernst. 2008. Visual-haptic cue weighting is independent of modality-specific attention. J Vis 8:21 21-16.

137. van Atteveldt, N. M., E. Formisano, R. Goebel, and L. Blomert. 2007. Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex. Neuroimage.

138. Fu, C. H. Y., A. R. McIntosh, J. Kim, W. Chau, E. T. Bullmore, S. C. R. Williams, G. D. Honey, and P. K. McGuire. 2006. Modulation of effective connectivity by cognitive demand in phonological verbal fluency. NeuroImage 30:266-271.

139. Gruber, O., T. Muller, and P. Falkai. 2007. Dynamic interactions between neural systems underlying different components of verbal working memory. Journal of Neural Transmission 114:1047-1050.

140. Husain, F. T., C. M. McKinney, and B. Horwitz. 2006. Frontal cortex functional connectivity changes during sound categorization. Neuroreport 17:617-621.

141. Noppeney, U., O. Josephs, J. Hocking, C. Price, and K. Friston. 2007. The effect of prior visual information on recognition of speech and sounds. Cereb Cortex 18:598-609.

142. Obleser, J., R. J. S. Wise, M. A. Dresner, and S. K. Scott. 2007. Functional integration across brain regions improves speech perception under adverse listening conditions. J Neurosci 27:2283-2289.

143. Patel, R. S., F. D. Bowman, and J. K. Rilling. 2006. Determining hierarchical functional networks from auditory stimuli fMRI. Human Brain Mapping 27:462-470.

144. Hampson, M., B. S. Peterson, P. Skudlarski, J. C. Gatenby, and J. C. Gore. 2002. Detection of functional connectivity using temporal correlations in MR images. Human Brain Mapping 15:247-262.

145. Noesselt, T., J. W. Rieger, M. A. Schoenfeld, M. Kanowski, H. Hinrichs, H. J. Heinze, and J. Driver. 2007. Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. J Neurosci 27:11431-11441.

146. Kreifelts, B., T. Ethofer, W. Grodd, M. Erb, and D. Wildgruber. 2007. Audiovisual integration of emotional signals in voice and face: an event-related fMRI study. NeuroImage 37:1445-1456.

147. Gentilucci, M., and L. Cattaneo. 2005. Automatic audiovisual integration in speech perception. Exp Brain Res 167:66-75.

148. Werner, S., and U. Noppeney. 2010. Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. Cereb Cortex 20:1829-1842.

149. Benoit, M. M., T. Raij, F.-H. Lin, I. P. Jaaskelainen, and S. Stufflebeam. 2010. Primary and multisensory cortical activity is correlated with audiovisual percepts. Human Brain Mapping 31:526-538.

150. Jones, J. A., and D. E. Callan. 2003. Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. Neuroreport 14:1129-1133.

151. Hertrich, I., S. Dietrich, and H. Ackerman. 2010. Cross-modal interactions during perception of audiovisual speech and nonspeech signals: an fMRI study. J Cogn Neurosci.

152. Wiersinga-Post, E., S. Tomaskovic, L. Slabu, R. Renken, F. de Smit, and H. Duihuis. 2010. Decreased BOLD responses in audiovisual processing. Neuroreport 00:000-000.

153. Colin, C., M. Radeau, and P. Deltenre. 2005. Top-down and bottom-up modulation of audiovisual integration in speech. European Journal of Cognitive Psychology 17:541-560.

154. Olson, I. R., J. C. Gatenby, and J. C. Gore. 2002. A comparison of bound and unbound audio-visual information processing in human cerebral cortex. Cognitive Brain Research 14:129-138.

155. Vul, E., C. Harris, P. Winkielman, and H. Pashler. 2009. Puzzlingly high correlations in fMRI studies of emotion, personality and social cognition. Perspectives on Psychological Science 4:274-290.

156. Saxe, R., M. Brett, and N. Kanwisher. 2006. Divide and conquer: a defense of functional localizers. Neuroimage 30:1088-1096; discussion 1097-1089.

157. Fischl, B., A. van der Kouwe, C. Destrieux, E. Halgren, F. Segonne, D. H. Salat, E. Busa, L. J. Seidman, J. Goldstein, D. Kennedy, V. Caviness, N. Makris, B. Rosen, and A. M. Dale. 2004. Automatically parcellating the human cerebral cortex. Cereb Cortex 14:11-22.

158. Hinds, O., S. Ghosh, T. W. Thompson, J. J. Yoo, S. Whitfield-Gabrieli, C. Triantafyllou, and J. D. Gabrieli. 2011. Computing moment-to-moment BOLD activation for real-time neurofeedback. Neuroimage 54:361-368.

159. Laconte, S. 2010. Decoding fMRI brain states in real-time. Neuroimage.

160. Hairston, W. D., J. H. Burdette, D. L. Flowers, F. B. Wood, and M. T. Wallace. 2005. Altered temporal profile of visual-auditory multisensory interactions in dyslexia. Exp Brain Res 166:474-480.

161. Blau, V. C., N. Van Atteveldt, M. Ekkebus, R. Goebel, and L. Blomert. 2009. Reduced neural integration of letters and speech sounds links phonological and reading deficits in adult dyslexia. Current Biology 19:503-508.

162. Rouger, J., B. Fraysse, O. Deguine, and P. Barone. 2008. McGurk effects in cochlear-implanted deaf subjects. Brain Res 1188:87-99.

163. Tremblay, C., F. Champoux, P. Voss, B. A. Bacon, F. Lepore, and H. Theoret. 2007. Speech and non-speech audio-visual illusions: a developmental study. PLoS ONE 2:e742.

164. Aziz-Zadeh, L., T. Sheng, and A. Gheytanchi. 2010. Common premotor regions for the perception and production of prosody and correlations with empathy and prosodic ability. PLoS One 5:e8759.

165. Champoux, F., C. Tremblay, M. Mercier, M. Lassonde, F. Lepore, J.-P. Gagne, and H. Theoret. 2006. A role for the inferior colliculus in multisensory speech integration. Neuroreport 17:1607-1610.

166. Hamilton, R. H., J. T. Shenton, and H. B. Coslett. 2006. An acquired deficit of audiovisual speech processing. Brain Lang 98:66-73.

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

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

169. Kushnerenko, E., T. Teinonen, A. Volein, and G. Csibra. 2008. Electrophysiological evidence of illusory audiovisual speech percept in human infants. Proc Nat Acad Sci U S A 105:11442-11445.

170. van Wassenhove, V., K. W. Grant, and D. Poeppel. 2005. Visual speech speeds up the neural processing of auditory speech. Proc Natl Acad Sci U S A 102:1181-1186.

171. Fingelkurts, A. A., A. A. Fingelkurts, C. M. Krause, R. Mottonen, and M. Sams. 2003. Cortical operational synchrony during audio-visual speech integration. Brain Lang 85:297-312.

172. Kaiser, J., I. Hertrich, H. Ackerman, K. Mathiak, and W. Lutzenberger. 2005. Hearing lips: Gamma-band activity during audiovisual speech perception. Cereb Cortex 15:646-653.

173. Mottonen, R., C. M. Krause, K. Tiippana, and M. Sams. 2002. Processing of changes in visual speech in the human auditory cortex. Cognitive Brain Research 13:417-425.

174. Surguladze, S. A., G. A. Calvert, M. J. Brammer, R. Campbell, E. T. Bullmore, V. Giampietro, and A. S. David. 2001. Audio-visual speech perception in schizophrenia: an fMRI study. Psychiatry Res 106:1-14.

175. Hasson, U., J. I. Skipper, H. C. Nusbaum, and S. L. Small. 2007. Abstract coding of audiovisual speech: beyond sensory representation. Neuron 56:1116-1126.

176. Beauchamp, M. S., K. E. Lee, and A. Martin. 2003. A region in posterior superior temporal sulcus that integrates auditory and visual information about complex objects. Neuroimage 19:S1428-S1428.

177. Van Essen, D. C. 2004. Surface-based approaches to spatial localization and registration in primate cerebral cortex. Neuroimage 23 Suppl 1:S97-S107.

178. Binder, J. R., J. A. Frost, T. A. Hammeke, P. S. Bellgowan, J. A. Springer, J. N. Kaufman, and E. T. Possing. 2000. Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex 10:512-528.

179. Huang, Q., and J. Tang. 2010. Age-related hearing loss or presbycusis. Eur Arch Otorhinolaryngol 267:1179-1191.

180. Klein, R., C. F. Chou, B. E. Klein, X. Zhang, S. M. Meuer, and J. B. Saaddine. 2011. Prevalence of age-related macular degeneration in the US population. Arch Ophthalmol 129:75-80.

181. Hockley, N., and L. Polka. 1994. A developmental study of audiovisual speech perception using the McGurk paradigm. J Acoust Soc Am 96:3309.

182. van Linden, S., and J. Vroomen. 2008. Audiovisual speech recalibration in children. Journal of Child Language 35:809-822.

183. Sekiyama, K., and D. Burnham. 2008. Impact of language on development of auditory-visual speech perception. Developmental Science 11:306-320.

184. Shaywitz, S. E., B. A. Shaywitz, J. M. Fletcher, and M. D. Escobar. 1990. Prevalence of reading disability in boys and girls. Results of the Connecticut Longitudinal Study. JAMA 264:998-1002.

185. Pekkola, J., M. Laasonen, V. Ojanen, T. Autti, I. P. Jaaskelainen, T. Kujala, and M. Sams. 2006. Perception of matching and conflicting audiovisual speech in dyslexic and fluent readers: an fMRI study at 3T. Neuroimage 29:797-807.

186. Bortfeld, H., E. Fava, and D. A. Boas. 2009. Identifying cortical lateralization of speech processing in infants using near-infrared spectroscopy. Dev Neuropsychol 34:52-65.

187. Frye, R. E., K. Hasan, L. Xue, D. Strickland, B. Malmberg, J. Liederman, and A. Papanicolaou. 2008. Splenium microstructure is related to two dimensions of reading skill. Neuroreport 19:1627-1631.

VITA

Audrey Rosa Nath was born in Houston, Texas on January 12, 1983, the Daughter of Rosa Chan Nath and Ravi Nath. After completing her work at Memorial High School, Houston, Texas in 2001, she entered Rice University in Houston, Texas. She received the degree of Bachelor of Science with a major in bioengineering as well as the Bachelor of Arts with Honors with a major in cognitive sciences from Rice in May, 2005. In June of 2005, she entered the MD/PhD Program at The University of Texas Medical School at Houston. She completed three years of medical school before beginning her PhD studies in Neuroscience at The University of Texas Health Science Center at Houston Graduate School of Biomedical Sciences.

Permanent address:
446 Mignon Ln.
Houston, TX 77024