Multisensory speech perception without the left superior temporal sulcus

Sarah H. Baum [a], Randi C. Martin [b], A. Cris Hamilton [b], Michael S. Beauchamp [a],*

[a] Department of Neurobiology and Anatomy, University of Texas Medical School at Houston, USA
[b] Department of Psychology, Rice University, USA

ARTICLE INFO

ABSTRACT

Converging evidence suggests that the left superior temporal sulcus (STS) is a critical site for multisensory integration of auditory and visual information during speech perception. We report a patient, SJ, who suffered a stroke that damaged the left tempo-parietal area, resulting in mild anomic aphasia. Structural MRI showed complete destruction of the left middle and posterior STS, as well as damage to adjacent areas in the temporal and parietal lobes. Surprisingly, SJ demonstrated preserved multisensory integration measured with two independent tests. First, she perceived the McGurk effect, an illusion that requires integration of auditory and visual speech. Second, her perception of morphed audiovisual speech with ambiguous auditory or visual information was significantly influenced by the opposing modality. To understand the neural basis for this preserved multisensory integration, blood-oxygen level dependent functional magnetic resonance imaging (BOLD fMRI) was used to examine brain responses to audiovisual speech in SJ and 23 healthy age-matched controls. In controls, bilateral STS activity was observed. In SJ, no activity was observed in the damaged left STS but in the right STS, more cortex was active in SJ than in any of the normal controls. Further, the amplitude of the BOLD response in right STS response to McGurk stimuli was significantly greater in SJ than in controls. The simplest explanation of these results is a reorganization of SJ's cortical language networks such that the right STS now subserves multisensory integration of speech.

Introduction

Speech can be understood through the auditory modality alone, but combining audition with vision improves speech perception (Grant and Seitz, 2000; Stein and Meredith, 1993; Sumby and Pollack, 1954). One striking behavioral example of audiovisual multisensory integration in speech perception is the McGurk effect (McGurk and MacDonald, 1976) in which an auditory syllable paired with a video clip of a different visual syllable results in the percept of a distinct new syllable (*e.g.* auditory "ba" + visual "ga" results in the percept "da"). Because the fused percept is different than either the auditory or visual stimulus, it can only be explained by multisensory integration.

A number of studies suggest that the left superior temporal sulcus (STS) is an important site of audiovisual multisensory integration. The left STS exhibits a larger BOLD response to multisensory stimuli as compared to unisensory stimuli (Beauchamp et al., 2004; Calvert et al., 2000; Stevenson and James, 2009). Tracer studies in rhesus macaque monkeys reveal that the STS is anatomically connected both

to auditory cortex and extrastriate visual cortex (Lewis and Van Essen, 2000; Seltzer et al., 1996). There is a correlation between the amplitude of activity in the left STS and the amount of McGurk perception in both individual adults (Nath and Beauchamp, 2012) and children (Nath et al., 2011). Inter-individual differences in left STS activity have also been linked to language comprehension abilities (McGettigan et al., 2012). When the left STS is temporarily inactivated with transcranial magnetic stimulation (TMS) in normal subjects, the McGurk effect is reduced (Beauchamp et al., 2010). Unlike the transient disruptions created by TMS, lesions caused by brain injury can give insight into the results of brain plasticity that occur after a stroke. In particular, damage to areas in the language network can result in brain reorganization, with increased activity in the areas homologous to the damaged tissue (Blasi et al., 2002; Buckner et al., 1996; Cao et al., 1999; Thomas et al., 1997; Winhuisen et al., 2005).

We describe a patient, SJ, with a lesion that completely ablated her left posterior STS. Following her stroke, SJ underwent intensive behavioral therapy. In the years following her stroke, her speech perception abilities improved. Five years after her stroke SJ demonstrated multisensory speech perception similar to 23 age-matched controls when tested with two independent behavioral measures. To understand the neural substrates of this ability, we examined patient SJ and age-matched controls with structural and functional MRI.

* Corresponding author at: 6431 Fannin St. Suite G.550, Houston, TX 77030, USA. Fax: +1 713 500 0623.
E-mail address: Michael.S.Beauchamp@uth.tmc.edu (M.S. Beauchamp).

Materials and methods

Patient SJ

All subjects provided informed 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 participants received compensation for their time. Patient SJ is a 63 year-old female who presented with a language impairment following a stroke, which destroyed a large portion of her left temporal lobe, including the left STS (Fig. 1 and Table 1). Patient SJ was 58 years old when she suffered a stroke in the left tempo-parietal area in September 2006. Prior to her stroke SJ worked in public relations and had completed one year of college. SJ's performance on the Western Aphasia Battery indicated a classification of anomic aphasia. Her auditory comprehension was impaired 3 years after the stroke (48% on auditory lexical decision and 86% for CV minimal pairs, compared with expected 95–100% for controls). 5 years after the stroke, her auditory recognition had improved to near normal range (87% on auditory lexical decision and 95% for CV minimal pairs). SJ was scanned two times, once for structural MRI in February 2010, and again for structural and functional MRI in March 2011.

Healthy age-matched control subjects

23 healthy older adults ranging in age from 53 to 75 years (14 female, mean age 62.9 years) served to provide a healthy age-matched comparison to patient SJ. Participants were recruited through word-of-mouth and flyers distributed around the greater Houston area. 21 subjects were right-handed as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971). All subjects were fluent English speakers.

Table 1

Anatomical regions impacted by stroke lesion. Column 1 shows the FreeSurfer automatic parcellation anatomical label. Column 2 shows the t-value of the volume difference between SJ and controls. All differences are statistically significant at a level of $p < 0.01$ corrected for multiple comparisons. Column 3 shows the difference between the gray matter volume in SJ and the average gray matter volume in 23 age-matched controls (column 4–column 5).

Label	t-value	Delta (mm^3)	Volume in SJ (mm^3)	Mean ± SD volume in controls (mm^3)
Supramarginal gyrus	6.8	−4984	438	5422 ± 714
Superior temporal sulcus	5.5	−4872	3038	7910 ± 867
Postcentral sulcus	4.1	−2023	1376	3399 ± 482
Inferior segment of the circular sulcus of the insula	4.7	−1853	547	2400 ± 385
Temporal plane of the superior temporal gyrus	4.4	−1732	4	1736 ± 389
Posterior segment of the lateral fissure	6.6	−1376	14	1390 ± 203
Anterior transverse temporal gyrus	4.8	−856	30	886 ± 174
Long insular gyrus and central sulcus of the insula	4.5	−747	287	1034 ± 163
Transverse temporal sulcus	4	−429	4	433 ± 287

Stimuli used for testing

Stimuli consisted of a digital video recording of a female native English speaker speaking "ba", "ga", "da", "pa", "ka" and "ta". Digital video editing software (iMovie, Apple Computer) was used to crop the total length of each video stimulus such that each clip both started and ended in a neutral, mouth-closed position. Each video clip ranged from 1.7 to 1.8 s.

Auditory-only stimuli were created by extracting the auditory track of each video and pairing it with white visual fixation crosshairs
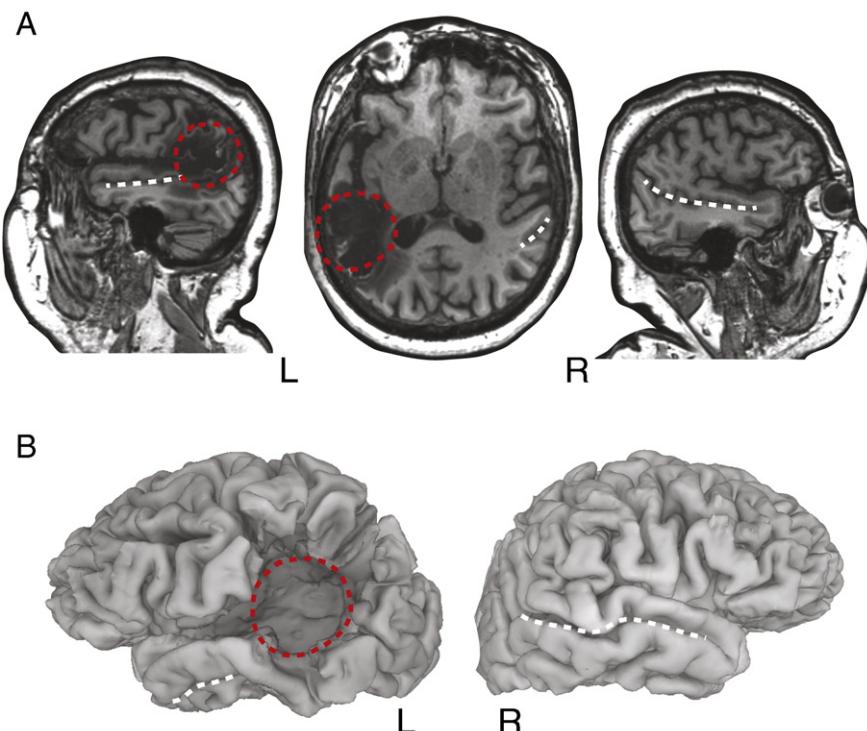

Fig. 1. Anatomical MRI of SJ. A. Sagittal and axial slices of SJ's structural MRI. White dashed lines indicate the location of the STS. Red dashed circle indicates stroke-damaged cortex in left hemisphere (left is left on all images). B. Cortical surface reconstruction of SJ's brain from the structural MRI. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

on a gray screen. Visual-only stimuli were created by removing the auditory track of each video. Two separate McGurk stimuli were created by pairing the auditory "ba" with the visual of "ga" (canonical percept "da"), and pairing the auditory "pa" with the visual of "ka" (canonical percept "ta"). Non-McGurk incongruent stimuli were created by reversing the pairing of the two McGurk stimuli (auditory "ga" with visual "ba", resulting in the percept "ga", and auditory "ka" with visual "pa", resulting in the percept "ka"). These stimuli were used for both behavioral testing and the fMRI experiment. Eight additional McGurk stimuli were obtained from youtube.com for additional behavioral testing with SJ.

Behavioral experiment

Behavioral testing of healthy controls

Each subject's perception of auditory only, congruent, and McGurk syllables was assessed. Stimuli were presented in two separate runs: auditory-only syllables (10 trials of each syllable) and AV syllables (10 trials each of "ba"/"da" McGurk syllables, "pa"/"ka" McGurk syllables, and "ba", "da", "pa" and "ka" congruent syllables) in random order. Auditory stimuli were delivered through headphones at approximately 70 dB, and visual stimuli were presented on a computer screen. For all stimuli, subjects were instructed to watch the mouth movements (if present) and listen to the speaker. Perception was assessed by asking the subject to verbally repeat out loud the perceived syllable. The response was open choice and no constraints were placed on allowed responses. This response format was chosen because it has been shown to provide a more conservative estimate of McGurk perception (Colin et al., 2005). All spoken responses were recorded by a microphone and the experimenter writing down each response. For SJ, the testing procedure was identical, but additional trials of McGurk stimuli were presented (15 trials vs. 10 in controls).

Morphed audiovisual syllables

An additional, independent, test of multisensory integration was obtained by measuring SJ's perception of audiovisual syllables along a continuum of "ba" to "da" (Massaro et al., 1993). Synthetic auditory speech stimuli were created by taking tokens of "ba" and "da" and manipulating the first 80 ms to create five auditory syllables ranging from A1 (100% ba/0% "da") to A5 (0% "ba"/100% "da"). Similarly, synthetic visible speech stimuli were created by using a computer-animated display whose mouth position at the syllable onset was systematically altered to create V1 (100% "ba"/0% "da") to V5 (0% "ba"/100% "da"). Each audiovisual syllable stimulus (five auditory times five visual for 25 total) was presented 20 times in random order in a two alternative forced choice task where SJ was instructed to respond if she perceived the audiovisual syllable to be more like "ba" or "da". Responses were made on a mouse with the left button labeled "ba" and the right button labeled "da". Written instructions were also presented on the screen after each trial. We compared SJ's responses with those of 82 healthy subjects viewing the same stimuli, reported in Massaro et al. (1993).

fMRI McGurk experiment

Each fMRI run lasted approximately 4 min and two scan runs were collected from each subject. In each run, single syllables were presented within the 2-second trial using a rapid event-related design. Trials were pseudo-randomized for an optimal rapid-event related order (Dale, 1999). In each trial, a video clip was presented followed by fixation crosshairs for the remainder of the trial. The crosshairs were positioned such that they were in the same location as the mouth during visual speech in order to minimize eye movements and draw attention to the visual mouth movements. Subjects responded to target trials only (the word "press"). For SJ and six control subjects, each run contained 25 McGurk trials, 25 non-McGurk incongruent trials, 25 congruent

trials, 20 target trials, and 25 trials of fixation baseline. For the remaining 17 control subjects each run contained 40 McGurk trials, 20 non-McGurk incongruent trials, 20 congruent trials, 15 target trials and 25 trials of fixation baseline. All stimuli were identical to those used for behavioral testing outside the scanner.

fMRI functional localizer experiment

In order to prevent bias when analyzing the McGurk fMRI data, a separate scan series was performed to identify independent regions of interest. The functional localizer scan consisted of six blocks of one syllable words (two auditory-only, two visual-only and two audiovisual blocks in random order) which contained 20 s of stimulus (10 two second trials, one word per trial) followed by 10 s of fixation baseline between each block. Each block contained a target trial (the word "press") of the same stimulus type as the other stimuli in the block; subjects were instructed to pay attention to each stimulus and press a button during target trials but not to any other stimuli.

MRI and fMRI analysis

Two T1-weighted MP-RAGE anatomical MRI scans were collected at the beginning of each scanning session with a 3 T whole-body MR scanner (Phillips Medical Systems) using a 32-channel head coil. The two anatomical scans were aligned to each other and averaged in order to provide maximal gray–white matter contrast. These scans were then used to create a cortical surface model using FreeSurfer (Dale et al., 1999; Fischl et al., 1999) for visualization in SUMA (Argall et al., 2006). For the fMRI scan series, T2* weighed images were collected using gradient echo-planar imaging (TR = 2000 ms, TE = 30 ms, flip angle = 90°) with in-plane resolution of 2.75 × 2.75 mm. The McGurk syllable scan series and localizer scan series consisted of 123 and 138 brain volumes, respectively. The first three volumes were discarded because they were collected before equilibrium magnetization was reached. This resulted in 120 and 135 usable brain volumes, respectively. Auditory stimuli were presented through MRI-compatible in-ear headphones (Sensimetrics, Malden, MA) which were covered with ear muffs to reduce the amount of noise from the scanner. Visual stimuli were presented on a projection screen with an LCD projector and viewed through a mirror attached to the head coil. Responses to the target trials were collected using a fiber-optic button response pad (Current Designs, Haverford, PA).

Analysis of the functional scan series was conducted using Analysis of Functional NeuroImages (AFNI) (Cox, 1996). Data were analyzed for each subject individually and then the data for all healthy control subjects was combined using a random-effects model. Functional data for each subject was first aligned to the averaged anatomical dataset and then motion-corrected using a local Pearson correlation (Saad et al., 2009). The analysis of all voxels was carried out with the AFNI function *3dDeconvolve,* which uses a generalized linear model utilizing a maximum-likelihood approach. Tent-zero functions were used in the deconvolution to estimate the individual hemodynamic response function in each voxel for each stimulus type that began at stimulus onset and ended 16 s after stimulus onset for rapid event related runs and 26 s for block design runs.

A modified, conservative t-test (Crawford and Howell, 1998) was used to compare single data points from SJ with averaged data from controls. To test for the significance of any differences in fMRI response amplitude by stimulus type, the within type variance was computed as follows. For controls, we considered the average response to a stimulus in each individual control subject as a sample. For SJ, we considered the response to individual presentations of each stimulus, calculated with a least-square sum method in the AFNI program *3dLSS* (Mumford et al., 2012). This analysis was used for all ROIs except for the left STS, for which the response was 0 for all trials, necessitating the use of the conservative single point t-test.

Group analysis

Two strategies were used for group analysis. Converging evidence from both strategies indicates a robust difference between SJ and controls. In the first strategy, regions of interest (ROI) are selected based on the individual anatomy in each subject (Saxe et al., 2006). Because the course of the STS is highly variable across subjects, standard 3-D anatomical templates fail to accurately align STS gray matter. Using a cortical-surface based analysis, the STS in each subject is aligned to the STS of a 2-D template for labeling purposes. This allows for unbiased measurement of activity in the STS (and other regions). Each ROI was created using the FreeSurfer anatomic parcellation of the cortical surface constructed from each individual subject's structural scans (Destrieux et al., 2010; Fischl et al., 2004). The parcellation defined 74 distinct regions for each hemisphere in each subject. SJ's automated parcellation was manually inspected to ensure that the 3-D reconstruction was an accurate representation of her structural damage. This parcellation was then manually edited for SJ's left hemisphere to ensure that no labels were assigned to the lesion zone.

ROIs created in each subject's individual native space were used in the main analysis, thus any potential discrepancy between the unnormalized brain and reference template did not affect the analysis results. These ROIs were then analyzed with data from independently collected runs, eliminating bias (Kriegeskorte et al., 2009). The STS ROI was defined by finding all voxels in the posterior half of the anatomically defined STS that responded to both auditory-only words and visual-only words (t>2 for each modality). For some subjects (n=5 in left hemisphere, n=1 in right hemisphere), there were no voxels in the posterior STS that were significantly active during both auditory-only and visual-only word blocks. For these subjects the STS ROI was defined by finding all voxels in the anatomically defined posterior STS that were active (t>2) during the audiovisual word blocks. The auditory cortex ROI was defined by finding voxels in the anatomically parcellated transverse temporal gyrus, lateral superior temporal gyrus and planum temporale that were active (t>2) during the auditory-only blocks. The extrastriate visual cortex ROI was defined by finding voxels in the anatomically parcellated extrastriate lateral occipitotemporal cortex that were active (t>2) during the visual-only blocks. We chose a later visual area to study because of its prominent role in visual speech perception and strong activation during audiovisual speech.

In the second strategy, a whole-brain voxel-wise analysis is used (Friston et al., 2006). Each individual subject brain and functional dataset was aligned to the N27 atlas brain (Mazziotta et al., 2001) with the *auto_tlrc* function in AFNI. The functional dataset for each subject was then smoothed using a $3 \times 3 \times 3$ mm FWHM Gaussian kernel. We wished to minimize blurring between the ROIs of interest and adjacent ROIs, so a small blurring kernel of approximately the same size as the voxel was chosen (Skudlarski et al., 1999). Areas with significantly different activation to McGurk stimuli between SJ and controls were searched for with *3dttest++*. These results were then transformed from the MRI volume to the cortical surface using *3dSurf2Vol* and clusters were identified with *SurfClust*. Clusters size threshold was 500 mm^2 with a z-score threshold of 3.5.

Results

Location and quantification of the lesion

Patient SJ's lesion destroyed a substantial portion of the lateral posterior left hemisphere (Fig. 1 and Table 1). To quantify the extent of the lesion, we used automated anatomical parcellation to compare SJ's left hemisphere with 23 age-matched controls. The supramarginal gyrus and the STS were the areas with the greatest loss of gray matter. The lesion also extended into the temporal plane of the superior temporal gyrus, the location of auditory cortex.

Auditory and McGurk perception: behavioral results

Sensory input is a prerequisite for multisensory integration. Because the lesion damaged regions of auditory cortex, we first examined SJ's auditory comprehension. When compared with 23 age-matched controls during our auditory-only syllable identification task, SJ was within the normal range (78% in SJ vs. 90% ± 15% in controls, $t_{22}=0.75$, $p=0.46$; Fig. 2A). Next, we examined SJ's perception of McGurk stimuli, incongruent auditory and visual syllables in which an illusory percept indicates the presence of multisensory integration. SJ and controls reported similar rates of the illusory McGurk percept (66% vs. 59% ± 42%, $t_{22}=0.16$, $p=0.87$; Fig. 2B).

Morphed audiovisual stimuli: behavioral results

As an independent test of multisensory integration, we presented 25 morphed audiovisual syllables along a continuum from "ba" to "da". SJ's perception was significantly influenced by both auditory and visual information. For instance, an ambiguous auditory stimulus (A4) was perceived as "da" 10% of the time when paired with one visual stimulus (V1) but was perceived as "da" 75% of the time when paired with a different visual stimulus (V5) ($p=10^{-8}$ with binomial distribution). Conversely, an ambiguous visual stimulus (V4) was perceived as "da" 35% when paired with one auditory stimulus (A1) but 75% when paired with a different auditory stimulus (A5) ($p=10^{-5}$ with binomial distribution). While SJ's multisensory integration in this task was significant, it was weaker for some stimuli than in the 82 controls tested by Massaro (1998) (A4V1, 10% vs. 66% ± 30% "da", $t_{81}=1.91$, $p=0.06$; A4V5, 75% vs. 98% ± 2%, $t_{81}=9.38$, $p=10^{-14}$; A1V4, 35% vs. 17% ± 25%, $t_{81}=0.69$, $p=0.49$; A5V4, 75% vs. 98% ± 2%, $t_{81}=8.62$, $p=10^{-13}$) (Fig. 2C).
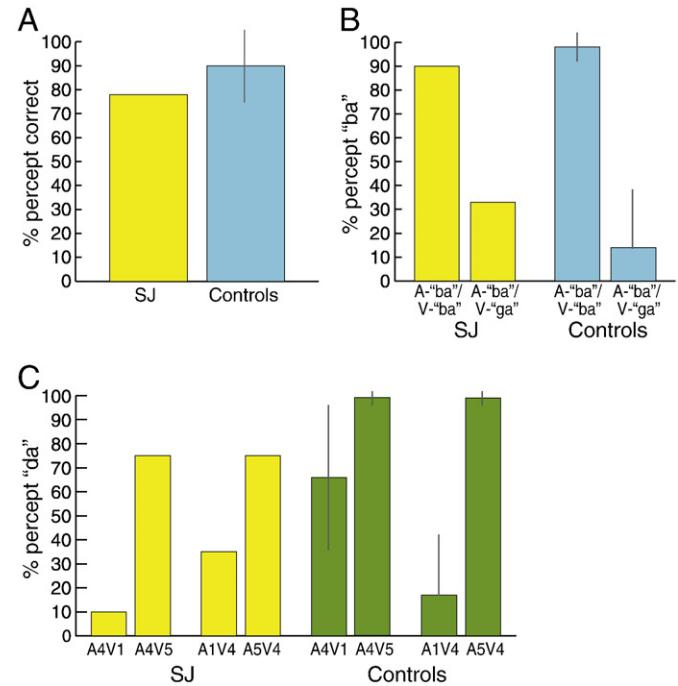

Fig. 2. Behavioral testing results. A. Averaged auditory-only performance for six syllables (chance performance 17%) for SJ (yellow) and age-matched controls (blue). B. Behavioral performance for one congruent audiovisual stimulus and one McGurk stimulus for SJ (yellow) and age-matched controls (blue). C. Behavioral performance with 4 exemplar audiovisual morphed syllables. Data for SJ (yellow) and controls (green); control data from (Massaro, 1998). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Functional MRI of patient SJ and controls

SJ's behavioral results showed evidence for multisensory integration despite the extensive damage to her left STS. To understand the neural substrates of this preserved integration, we used fMRI to examine brain responses to multisensory speech.

We first presented separate blocks of auditory, visual and audiovisual words. Normal controls showed bilateral responses to audiovisual speech stimuli, with especially strong responses in the left superior temporal gyrus (STG) and STS. As expected from the extensive lesional damage, no activity was observed in SJ's left STS. However, activity was observed in her right hemisphere. Especially for the right STS, this activity appeared more extensive than in normal controls (Fig. 3A). We used three strategies to quantify this observation. First, we measured the volume of active cortex within ROIs as defined by the localizer scan consisting of whole words. Second, we measured the amplitude of the response within localizer-defined ROIs to McGurk stimuli. Third, we performed a whole-brain analysis of activity evoked by the McGurk stimuli.

Method 1: volume of activated cortex

To quantify activity, we measured the volume of cortex that showed significant responses to whole word audiovisual speech in three regions of interest: the STS, lateral extrastriate visual cortex, and auditory cortex (Fig. 3B). As expected from the damage caused by the lesion, there was no active cortex in SJ's left STS vs. a large volume of STS activation in controls (0 vs. 34 ± 27 mm^3, $t_{22} = 6.18$, $p = 10^{-6}$) (Fig. 4A). However, in right STS, SJ had much *more* active cortex than normal controls (96 vs. 30 ± 20 mm^3, $t_{22} = 3.21$, $p = 0.004$). In fact, the volume of active cortex in SJ's right STS was greater than in any normal individual (Fig. 4B). This finding (less active cortex in left hemisphere, more active cortex in right hemisphere) was not found in other ROIs. In extrastriate visual cortex, located close to the STS but just posterior and ventral to the lesion zone, there was no significant difference between SJ and controls in either the left hemisphere (174 vs. 152 ± 68 mm^3, $t_{22} = 0.32$, $p = 0.75$) or the right hemisphere (164 vs. 167 ± 70 mm^3, $t_{22} = 0.04$, $p = 0.97$). In auditory cortex, which overlapped the lesion zone,

there was less active cortex in left hemisphere in SJ compared with controls (75 vs. 242 ± 76 mm^3, $t_{22} = 2.16$, $p = 0.04$) and no difference in right hemisphere (202 vs. 213 ± 71 mm^3, $t_{22} = 0.15$, $p = 0.88$).

Method 2: amplitude of HDR to McGurk stimuli

Next, we examined the amplitude of the response to McGurk stimuli within the STS, visual cortex, and auditory cortex ROIs. Because these ROIs were created with independent localizer scans that contained words and not McGurk stimuli, the analysis was not biased (Kriegeskorte et al., 2009; Vul et al., 2009). There was no response in SJ's left STS (0% in SJ vs. 0.11% in controls $t_{22} = 4.25$, $p = 10^{-4}$) but the response in SJ's right STS was significantly greater that controls (0.29% in SJ vs 0.13% in controls, $t_{71} = 2.57$, $p = 0.01$) (Fig. 4C). This pattern (less activity than controls in left hemisphere, more activity than controls in right hemisphere) was not found in other ROIs. In visual cortex, there were no significant difference in McGurk amplitude in the left extrastriate cortex (0.07% in SJ vs 0.10% in controls, $t_{71} = 0.67$, $p = 0.50$) while right hemisphere showed greater response (0.21% in SJ vs 0.12% in controls, $t_{71} = 1.96$, $p = 0.05$). In auditory cortex, SJ's response was significantly weaker in left hemisphere (-0.06% in SJ vs 0.22% in controls, $t_{71} = 5.64$, $p = 3 \times 10^{-7}$) but was similar to controls in right hemisphere (0.26% in SJ vs 0.19% in controls, $t_{71} = 1.33$, $p = 0.19$).

If SJ's right STS subserved new functions because of the lesion to SJ's left STS, we would expect a differential pattern of activity in SJ's right STS compared to other right hemisphere ROIs. To test this idea, we performed an ANOVA on right hemisphere responses to McGurk stimuli across the ROIs between SJ and controls (the variance was computed within subject for SJ and across subjects for controls). A main effect of subject group (SJ vs. controls) would suggest that all right hemisphere ROIs showed different responses between SJ and controls. A main effect of ROI (STS, auditory cortex, visual cortex) would suggest that a particular ROI was more active, regardless of group. A significant interaction would suggest differential effects between different right hemisphere ROIs between SJ and controls. The ANOVA found a significant interaction between group and ROI ($F_{2,213} = 4.70$, $p = 0.01$) without significant main effects for group or ROI. This suggests that the different ROIs in the right hemisphere
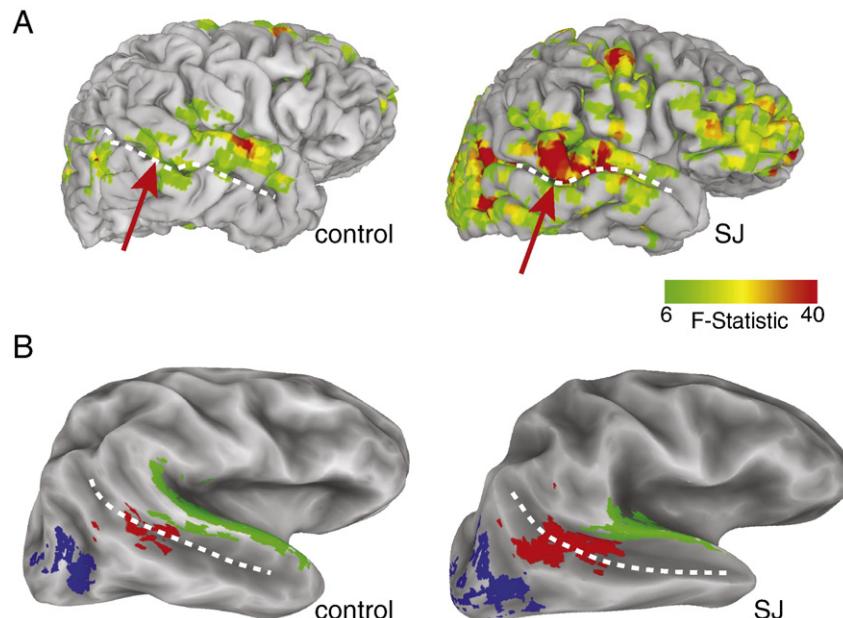

Fig. 3. fMRI activation during localizer scan. A. Response to audiovisual speech in right hemisphere (lateral view of cortical surface model, color scale indicates significance of response) in one age-matched control (left, case IN) and stroke patient (right, case SJ). White dashed lines indicate STS, red arrow indicates activity in right STS. B. Location of STS (red), extrastriate visual cortex (blue), and auditory cortex (green) ROIs in the right hemisphere of age-matched control (left, case IN) and stroke patient (right, case SJ). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)
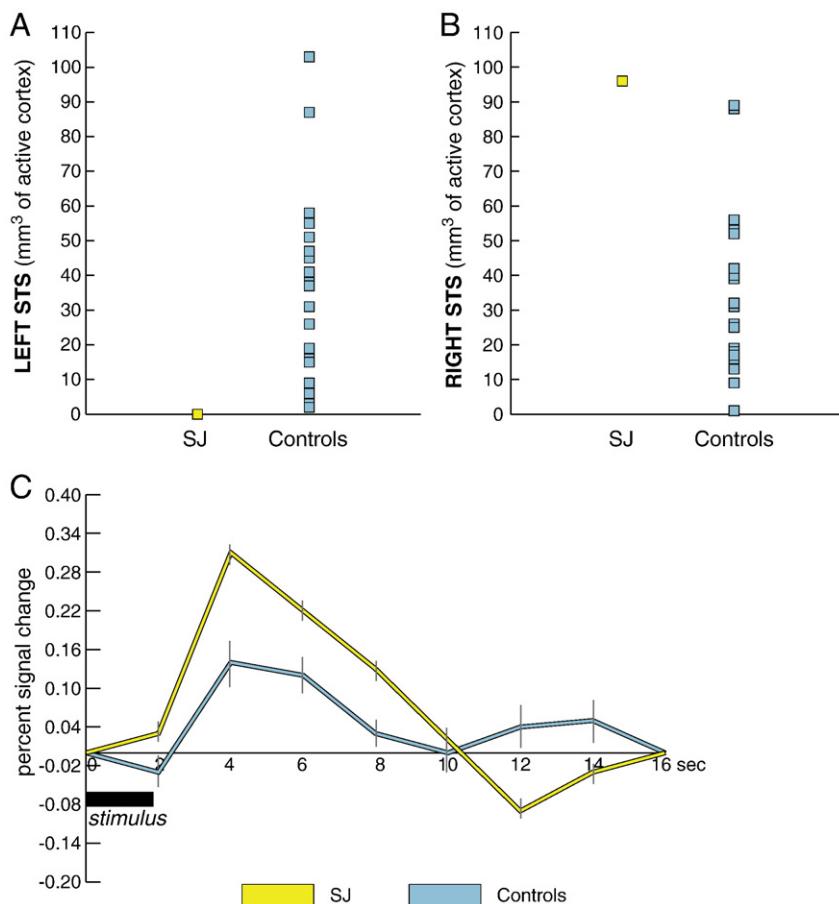
Fig. 4. Multisensory responses in the STS in SJ and controls. A. Volume of active cortex in the left STS of SJ (yellow) and age-matched controls (blue). B. Volume of active cortex in the right STS of SJ (yellow) and age-matched controls (blue). C. Hemodynamic response for SJ (yellow) and healthy controls (blue) in the right STS in response to the McGurk syllable A-"ba"/V-"ga". Error bars denote standard error of the mean (within-subject variance for SJ and between-subject variance for controls). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

responded differently in SJ compared with controls, driven by a greater a response in right STS in SJ compared with controls.

Method 3: whole brain analysis

In a third strategy to look for neural differences between SJ and controls, we performed a whole brain analysis of the response to McGurk stimuli. Regions with both increased and decreased responses relative to controls were observed (Table 2). The region with the largest area of increased activity in SJ relative to controls was in the right STS. The region with the largest decrease in activity in SJ relative to controls was in the left STS and the remainder of the lesion zone in the left hemisphere.

Amplitude of HDR to congruent and non-McGurk incongruent stimuli

In addition to McGurk stimuli (which were of greatest interest because they require multisensory integration) we also measured the response to congruent stimuli and non-McGurk incongruent stimuli. In the STS of normal controls, the largest response was to non-McGurk incongruent stimuli with significantly weaker responses to congruent and McGurk stimuli (incongruent stimuli: 0.22% in left STS, 0.25% in right STS compared with congruent: 0.16% in left STS, $t_{22} = 2.74$, $p = 0.01$; 0.17% in right STS, $t_{22} = 3.08$, $p = 0.01$; compared with McGurk: 0.14% in left STS, $t_{22} = 2.41$, $p = 0.03$; 0.14% in right STS, $t_{22} = 3.08$, $p = 0.01$; no significant hemispheric differences) (Fig. 5A). This response pattern was markedly altered in SJ. Instead of the maximal response to non-McGurk incongruent stimuli observed in controls, SJ had similar

amplitudes of response to each stimulus type in her right STS (non-McGurk incongruent = 0.25%, McGurk = 0.29%, congruent = 0.29%, $F_{2,147} = 0.33$, $p = 0.72$) (Fig. 5B).

Discussion

We examined a subject, SJ, whose stroke completely destroyed a large portion of her left temporal lobe, including the left STS. Previous studies have demonstrated a critical role of the left STS in multisensory speech perception (Beauchamp, 2005; Miller and D'Esposito, 2005; Nath and Beauchamp, 2011, 2012; Scott and Johnsrude, 2003;

Table 2
Areas of differential activation in SJ and controls. Regions from the whole brain analysis of significant difference in response to McGurk stimuli between SJ and age-matched controls, mapped to the cortical surface. Regions are ranked by area on the cortical surface. Talairach coordinates following anatomical label in *(x, y, z)* format are the weighted center of mass of the cluster.

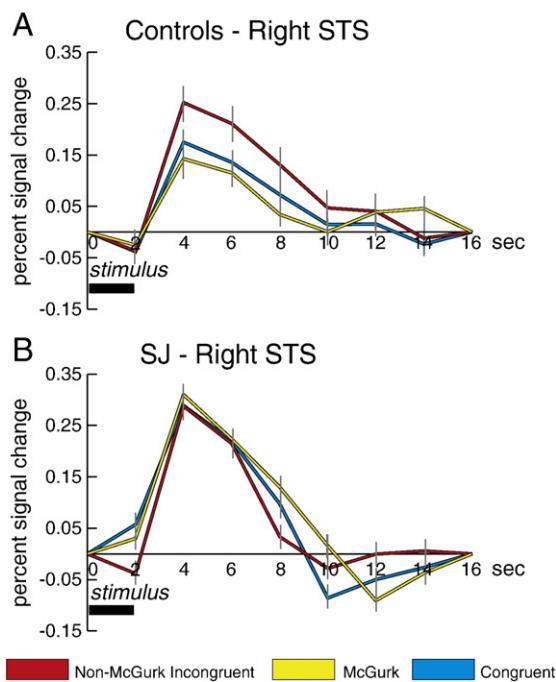| Label	Area (mm^2)
Increased activity in SJ	
R superior temporal sulcus *(45, −59, 18)*	1250
R superior frontal sulcus *(29, 9, 48)*	1120
L frontomarginal sulcus *(−29, 49, 2)*	935
L central sulcus *(−25, −31, 62)*	748
L angular gyrus *(−45, −69, 22)*	731
R frontomarginal sulcus *(33, 49, −2)*	547
Decreased activity in SJ	
L lateral-posterior temporal, including STS *(−43, −24, 7)*	4449
L postcentral sulcus *(−35, −43, 38)*	695

Fig. 5. Hemodynamic response to all audiovisual stimuli in SJ and controls. A. Response to non-McGurk incongruent (red), McGurk (yellow) and congruent (blue) audiovisual stimuli in the right STS of age-matched controls. Error bars denote standard error of the mean across subjects. B. Response to the same stimuli in the right STS of SJ. Error bars denote standard error of the mean within SJ. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Stevenson and James, 2009). Because temporary disruption of the left STS with TMS impairs multisensory speech perception (Beauchamp et al., 2010) one might expect the lesion suffered by SJ to greatly reduce multisensory integration. Surprisingly, patient SJ showed robust multisensory integration when tested with two independent behavioral tests 5 years after her stroke.

Evidence suggests that SJ's speech perception abilities changed in the years following her stroke, during which she received extensive rehabilitation therapy. She spent 12 h a week for approximately 40 weeks a year in the years following her stroke at the Houston Aphasia Recovery Center as well as receiving additional speech and language therapy. SJ and her husband report that this intensive therapy has been extremely beneficial to her recovery. Consistent with this anecdotal report, SJ's speech perception abilities improved following her stroke, from 48% on auditory lexical decision 3 years following the stroke to 87% at 5 years following the stroke (because multisensory integration was only tested 5 years following the stroke, we do not know whether SJ's multisensory abilities showed a parallel improvement.)

Based on the observed improvements in speech perception, neural plasticity and rehabilitation in SJ might have resulted in brain changes, leading to her improved abilities. This would predict different patterns of brain activity during multisensory speech perception in SJ compared with age-matched controls. To test this hypothesis, we studied the neuroanatomical substrates of multisensory speech perception with structural and functional MRI in SJ and 23 age-matched controls. Age-matched controls had large volumes of active multisensory cortex in both the left and right STS when perceiving audiovisual speech. In comparison, speech evoked no activity in SJ's left STS but a larger volume of active cortex in right STS than in any age-matched control. The response amplitude to McGurk stimuli in the right STS was significantly greater than the right STS response in the healthy age-matched controls. These results suggest that SJ's multisensory speech perception may be supported by her right STS. As auditory noise increases, multisensory integration becomes more important (Ross et al., 2007). SJ's diminished auditory abilities

immediately following her stroke may have driven the recruitment of right hemisphere areas in the service of multisensory integration for speech comprehension.

A notable finding is that the response amplitude in SJ's right STS to all three types of audiovisual syllables was large and relatively uniform, in contrast with the maximal activation to incongruent stimuli observed in healthy controls (Stevenson et al., 2011; van Atteveldt et al., 2010). This could reflect an attentional effect, in which healthy subjects automatically process most audiovisual speech, with an enhanced response to incongruent stimuli because they attract attention. SJ's right STS processing of speech may require more conscious effort on her part, resulting in attentional modulation (and enhanced response) for all audiovisual speech stimuli. Indeed, SJ reports that watching speakers on TV (such as a newscast) or conversing with others is especially mentally effortful.

Our results are consistent with a large body of literature showing that the contralesional hemisphere is able to compensate for damage after a brain injury. Left hemisphere strokes often result in aphasia (Dronkers et al., 2004) that resolves (at least partially) over time. Functional imaging studies of these cases have demonstrated increased activity in right-hemisphere homologues of left hemisphere language areas (Blasi et al., 2002; Buckner et al., 1996; Cao et al., 1999; Thomas et al., 1997; Winhuisen et al., 2005). While these studies used high-level language tasks, such as word retrieval, we observed similar right hemisphere compensation in a low-level task that required integration of auditory and visual speech information.

While the finding that SJ has multisensory integration is surprising based on the McGurk perception literature from healthy controls, it is in line with other reports from aphasics in the literature showing that aphasics are able to integrate sensory information. Champoux et al. (2006) examined a child with damage to the right inferior colliculus and noted that when McGurk stimuli were presented in the left hemifield, the patient's perception of the illusion was dramatically reduced. McGurk fusion percepts have also been found in stroke patients whose lesion locations are less well defined (Campbell et al., 1990; Schmid et al., 2009). Youse et al. (2004) describe a patient, JP, who suffered a left hemisphere stroke and perceived the McGurk effect (although poor performance on the auditory-only syllables makes this more difficult to interpret than in SJ). Other audiovisual integration effects have been noted in patients who presented with visual neglect, hemianopia, or both (Frassinetti et al., 2005). An important distinction is between auditory-visual language stimuli in which both modalities are presented in their natural speech form (i.e. auditory "ba" + video of speaker saying "ba") with an orthographic representation (i.e. auditory "ba" + printed letters "ba"). Although orthographic auditory–visual tasks also recruit the STS (Blau et al., 2008; Raij et al., 2000; van Atteveldt et al., 2004) there are differences between letter-speech and audiovisual speech processing (Froyen et al., 2010) and lesions might be expected to differentially impair these two tasks. For instance, Hickok et al. (2011) found that Broca's aphasics were impaired on an auditory–visual grapheme discrimination task.

We observed significant variability within our population of 23 age-matched controls, which may be linked to individual differences in multisensory integration and language ability (Kherif et al., 2009; McGettigan et al., 2012; Nath and Beauchamp, 2012; Nath et al., 2011). Because we do not have pre-injury data for SJ, we cannot refute the null hypothesis that her right hemisphere subserved multisensory integration even before the stroke and that no cortical reorganization occurred. However, the observation that SJ's volume of speech-evoked activity in right STS was greater than in any age-matched control (and that no activity was observed in SJ's left STS, far less than in any age-matched control) supports a neural plasticity explanation. SJ's extensive rehabilitation efforts are similar to those known to cause dramatic reorganization in language networks, such as in illiterate adults undergoing literacy training (Carreiras et al., 2009).

While our study does not provide direct evidence that the activity observed in SJ's right STS is critical for her multisensory abilities, other studies have shown that disrupting the right hemisphere of recovered aphasia patients using TMS (Winhuisen et al., 2005), intracarotid amobarbital (Czopf, 1979; Kinsbourne, 1971) or even additional infarcts (Turkeltaub et al., 2011) results in profound language impairments. We hypothesize that a similar manipulation, such as TMS of SJ's right STS, would greatly reduce her multisensory speech perception.

Acknowledgments

References

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

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

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

Beauchamp, M.S., Nath, A.R., Pasalar, S., 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.

Blasi, V., Young, A.C., Tansy, A.P., Petersen, S.E., Snyder, A.Z., Corbetta, M., 2002. Word retrieval learning modulates right frontal cortex in patients with left frontal damage. Neuron 36, 159–170.

Blau, V., van Atteveldt, N., Formisano, E., Goebel, R., Blomert, L., 2008. Task-irrelevant visual letters interact with the processing of speech sounds in heteromodal and unimodal cortex. Eur. J. Neurosci. 28, 500–509.

Buckner, R.L., Corbetta, M., Schatz, J., Raichle, M.E., Petersen, S.E., 1996. Preserved speech abilities and compensation following prefrontal damage. Proc. Natl. Acad. Sci. U.S.A. 93, 1249–1253.

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

Campbell, R., Garwood, J., Franklin, S., Howard, D., Landis, T., Regard, M., 1990. Neuropsychological studies of auditory–visual fusion illusions. Four case studies and their implications. Neuropsychologia 28, 787–802.

Cao, Y., Vikingstad, E.M., George, K.P., Johnson, A.F., Welch, K.M., 1999. Cortical language activation in stroke patients recovering from aphasia with functional MRI. Stroke 30, 2331–2340.

Carreiras, M., Seghier, M.L., Baquero, S., Estevez, A., Lozano, A., Devlin, J.T., Price, C.J., 2009. An anatomical signature for literacy. Nature 461, 983–986.

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

Colin, C., Radeau, M., Deltenre, P., 2005. Top-down and bottom-up modulation of audiovisual integration in speech. Eur. J. Cogn. Psychol. 17, 541–560.

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

Crawford, J.R., Howell, D.C., 1998. Comparing an individual's test scores against norms derived from small samples. Clin. Neuropsychol. 12, 482–486.

Czopf, D., 1979. The role of the non-dominant hemisphere in speech recovery. Aphasia Apraxia Agnosia 2, 27–33.

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

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

Destrieux, C., Fischl, B., Dale, A., Halgren, E., 2010. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Neuroimage 53, 1–15.

Dronkers, N.F., Wilkins, D.P., Van Valin Jr., R.D., Redfern, B.B., Jaeger, J.J., 2004. Lesion analysis of the brain areas involved in language comprehension. Cognition 92, 145–177.

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

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

Frassinetti, F., Bolognini, N., Bottari, D., Bonora, A., Ladavas, E., 2005. Audiovisual integration in patients with visual deficit. J. Cogn. Neurosci. 17, 1442–1452.

Friston, K.J., Rotshtein, P., Geng, J.J., Sterzer, P., Henson, R.N., 2006. A critique of functional localisers. Neuroimage 30, 1077–1087.

Froyen, D., van Atteveldt, N., Blomert, L., 2010. Exploring the role of low level visual processing in letter-speech sound integration: a visual MMN study. Front. Integr. Neurosci. 4, 9.

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

Hickok, G., Costanzo, M., Capasso, R., Miceli, G., 2011. The role of Broca's area in speech perception: evidence from aphasia revisited. Brain Lang. 119, 214–220.

Kherif, F., Josse, G., Seghier, M.L., Price, C.J., 2009. The main sources of intersubject variability in neuronal activation for reading aloud. J. Cogn. Neurosci. 21, 654–668.

Kinsbourne, M., 1971. The minor cerebral hemisphere as a source of aphasic speech. Arch. Neurol. 25, 302–306.

Kriegeskorte, N., Simmon, W.K., Bellgowan, P.S., Baker, C.I., 2009. Circular analysis in systems neuroscience: the dangers of double dipping. Nat. Neurosci. 12, 535–540.

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

Massaro, D., 1998. Data archive of 5 5 + 5 + 5 expanded factorial visual–auditory recognition experiments. (URL: http://mambo.ucsc.edu/psl/8236/).

Massaro, D., Cohen, M.M., Gesi, A., Heredia, R., Tsuzaki, M., 1993. Bimodal speech perception: an examination across languages. J. Phon. 21, 445–478.

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

McGettigan, C., Faulkner, A., Altarelli, I., Obleser, J., Baverstock, H., Scott, S.K., 2012. Speech comprehension aided by multiple modalities: behavioural and neural interactions. Neuropsychologia 50 (5), 762–776 (April).

McGurk, H., MacDonald, J., 1976. Hearing lips and seeing voices. Nature 264, 746–748.

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

Mumford, J.A., Turner, B.O., Ashby, F.G., Poldrack, R.A., 2012. Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. Neuroimage 59, 2636–2643.

Nath, A.R., Beauchamp, M.S., 2011. Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. J. Neurosci. 31, 1704–1714.

Nath, A.R., Beauchamp, M.S., 2012. A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. Neuroimage 59, 781–787.

Nath, A.R., Fava, E.E., Beauchamp, M.S., 2011. Neural correlates of interindividual differences in children's audiovisual speech perception. J. Neurosci. 31, 13963–13971.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the edinburgh inventory. Neuropsychologia 9, 97–113.

Raij, T., Uutela, K., Hari, R., 2000. Audiovisual integration of letters in the human brain. Neuron 28, 617–625.

Ross, L.A., Saint-Amour, D., Leavitt, V.M., Javitt, D.C., Foxe, J.J., 2007. Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. Cereb. Cortex 17, 1147–1153.

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

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

Schmid, G., Thielmann, A., Ziegler, W., 2009. The influence of visual and auditory information on the perception of speech and non-speech oral movements in patients with left hemisphere lesions. Clin. Linguist. Phon. 23, 208–221.

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

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

Skudlarski, P., Constable, R.T., Gore, J.C., 1999. ROC analysis of statistical methods used in functional MRI: individual subjects. Neuroimage 9, 311–329.

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

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

Stevenson, R.A., VanDerKlok, R.M., Pisoni, D.B., James, T.W., 2011. Discrete neural substrates underlie complementary audiovisual speech integration processes. Neuroimage 55, 1339–1345.

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

Thomas, C., Altenmueller, E., Marckmann, G., Kahrs, J., Dichgans, J., 1997. Language processing in aphasia: changes in lateralization patterns during recovery reflect cerebral plasticity in adults. Electroencephalogr. Clin. Neurophysiol. 102, 86–97.

Turkeltaub, P.E., Coslett, H.B., Thomas, A.L., Faseyitan, O., Benson, J., Norise, C., Hamilton, R.H., 2011. The right hemisphere is not unitary in its role in aphasia recovery. Cortex. http://dx.doi.org/10.1016/j.cortex.2011.06.010 (epub ahead of print).

van Atteveldt, N., Formisano, E., Goebel, R., Blomert, L., 2004. Integration of letters and speech sounds in the human brain. Neuron 43, 271–282.

van Atteveldt, N.M., Blau, V.C., Blomert, L., Goebel, R., 2010. fMR-adaptation indicates selectivity to audiovisual content congruency in distributed clusters in human superior temporal cortex. BMC Neurosci. 11, 11.

Vul, E., Harris, C., Winkielman, P., Pashler, H., 2009. Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. Perspect. Psychol. Sci. 4, 274–290.

Winhuisen, L., Thiel, A., Schumacher, B., Kessler, J., Rudolf, J., Haupt, W.F., Heiss, W.D., 2005. Role of the contralateral inferior frontal gyrus in recovery of language function in poststroke aphasia: a combined repetitive transcranial magnetic stimulation and positron emission tomography study. Stroke 36, 1759–1763.

Youse, K.M., Cienkowski, K.M., Coelho, C.A., 2004. Auditory–visual speech perception in an adult with aphasia. Brain Inj. 18, 825–834.