Behavioral/Systems/Cognitive

Neural Correlates of Interindividual Differences in Children's Audiovisual Speech Perception

Audrey R. Nath, Eswen E. Fava, and Michael S. Beauchamp
Department of Neurobiology and Anatomy, University of Texas Medical School at Houston, Houston, Texas 77030

Children use information from both the auditory and visual modalities to aid in understanding speech. A dramatic illustration of this multisensory integration is the McGurk effect, an illusion in which an auditory syllable is perceived differently when it is paired with an incongruent mouth movement. However, there are significant interindividual differences in McGurk perception: some children never perceive the illusion, while others always do. Because converging evidence suggests that the posterior superior temporal sulcus (STS) is a critical site for multisensory integration, we hypothesized that activity within the STS would predict susceptibility to the McGurk effect. To test this idea, we used BOLD fMRI in 17 children aged 6–12 years to measure brain responses to the following three audiovisual stimulus categories: McGurk incongruent, non-McGurk incongruent, and congruent syllables. Two separate analysis approaches, one using independent functional localizers and another using whole-brain voxel-based regression, showed differences in the left STS between perceivers and nonperceivers. The STS of McGurk perceivers responded significantly more than that of nonperceivers to McGurk syllables, but not to other stimuli, and perceivers' hemodynamic responses in the STS were significantly prolonged. In addition to the STS, weaker differences between perceivers and nonperceivers were observed in the fusiform face area and extrastriate visual cortex. These results suggest that the STS is an important source of interindividual variability in children's audiovisual speech perception.

Introduction

From infancy, children use the independent information available from the auditory modality (heard speech) and the visual modality (mouth movements) when perceiving speech (Massaro et al., 1986; Sekiyama and Burnham, 2008; van Linden and Vroomen, 2008). This integration of auditory and visual speech streams can be demonstrated by a remarkable illusion known as the McGurk effect (McGurk and MacDonald, 1976): an auditory "ba" presented with the mouth movements of "ga" is perceived by the listener as a completely different syllable, "da" (the McGurk percept). The McGurk effect has been used in behavioral studies to probe the development of audiovisual speech, demonstrating that some infants are able to perceive a McGurk-like percept as early as 4–5 months of age (Rosenblum et al., 1997; Burnham and Dodd, 2004). However, there is substantial interindividual variability in children's perception of the McGurk effect (McGurk and MacDonald, 1976; Dupont et al., 2005; Tremblay et al., 2007). A behavioral study found that only 57% of children 5–14 years old perceived the McGurk effect in at least 70% of trials (Schorr et al., 2005). The goal of our study was to examine neural responses to McGurk stimuli in perceiving and nonperceiving children to provide a better understanding of the neural basis for interindividual differences in audiovisual speech perception in children.

Studies in adults have suggested that the superior temporal sulcus (STS) is a critical region for multisensory integration of audiovisual speech (Calvert et al., 2000; Sekiyama et al., 2003; Callan et al., 2004; Miller and D'Esposito, 2005; Stevenson and James, 2009) and is especially important for the McGurk effect (Beauchamp et al., 2010). The STS is also likely to be important for audiovisual integration in children, because it is active during presentation of audiovisual speech (Dick et al., 2010). Increased activity in the STS is a neural signature for multisensory integration (Wright et al., 2003; Beauchamp et al., 2004; Van Atteveldt et al., 2004). Therefore, we predicted that the STS should be active in McGurk perceivers (reflecting their integration of the auditory and visual components of the McGurk stimulus) but not in nonperceivers (reflecting their lack of audiovisual integration). Because the STS is only one node in the language network, we also examined other areas important for language processing in children, including auditory cortex, the fusiform gyrus, and extrastriate visual cortex (MacSweeney et al., 2002; Schlagger et al., 2002; Devlin et al., 2006; Cone et al., 2008; Cao et al., 2010; Dick et al., 2010; Lidzba et al., 2011).

Materials and Methods

Subjects and stimuli. Seventeen healthy children ranging in age from 6 to 12 years (10 female, mean age 9.3 ± 2.2 years) (Table 1 for demographics) participated in the study. Two additional subjects were excluded: one for excessive head motion (>40 mm) and another who was unable to hear the stimuli after the headphones became dislodged during the experiment. The subjects were right handed according to the Edinburgh handedness inventory (Oldfield, 1971), native English speakers, and re-

Received May 25, 2011; revised July 19, 2011; accepted Aug. 8, 2011.

Author contributions: A.R.N. and M.S.B. designed research; A.R.N., E.E.F., and M.S.B. performed research; A.R.N. and M.S.B. contributed unpublished reagents/analytic tools; A.R.N. and M.S.B. analyzed data; A.R.N. and M.S.B. wrote the paper.

This research was supported by National Science Foundation Grant 642532, and NIH Grants R01NS065395, TL1RR024147, S10RR019186, and F31DC009765. We thank Vips Patel for assistance with MR data collection.

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

DOI:10.1523/JNEUROSCI.2605-11.2011

 0270-6474/11/3113963-09$15.00/0

ported no hearing or vision impairments. All subjects provided assent, and informed consent from a parent was obtained 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" (McGurk and MacDonald, 1976). Digital video editing software (iMovie, Apple) was used to crop the original recordings. The duration of the auditory syllables ranged from 0.4 to 0.5 s. The total length of each video clip ranged from 1.7 to 1.8 s to start and end each video in a neutral, mouth-closed position and to include all mouth movements from mouth opening to closing. Congruent audiovisual stimuli consisted of synchronous audiovisual recordings of "ba," "da," and "ma." Auditory-only syllables consisted of the auditory components of "ba" or "da" presented with a white visual fixation crosshairs.

The following three types of audiovisual stimuli were created: McGurk incongruent, non-McGurk incongruent, and congruent syllables (Fig. 1*A–C*). We created the McGurk syllable (auditory "ba" + visual "ga"), producing the McGurk percept of "da," and a non-McGurk incongruent syllable ("ga" + visual "ba"), producing an auditory percept such as "ga" or a combination percept such as "g-ba." The congruent syllable consisted of synchronous auditory "ba" and visual "ba."

The functional localizer consisted of unisensory auditory and visual stimuli. Auditory-only words were used because they reliably activate auditory cortex (Belin et al., 2002; Poeppel et al., 2004) without requiring complex semantic and syntactic processing. The auditory stimuli were drawn from 200 single-syllable words from the Medical Research Council Psycholinguistic Database with Brown verbal frequency of 20–200, imageability rating of >100, age of acquisition of <7 years, and Kucera-Francis written frequency >80 (Wilson, 1988). In pilot studies of unisensory visual stimuli, silent videos of visually mouthed words were tested, but children were puzzled by the absence of speech sounds. Therefore, we substituted videos of silent facial emotion, salient stimuli that are reliable activators of visual areas (Ishai et al., 2005; Kessler et al., 2011; Sabatinelli et al., 2011) and do not have strong auditory associations. Each silent video was 2.5 s in length and contained one of 11 models emoting one of four primary facial expressions (anger, happiness, surprise, and sadness).

Behavioral McGurk experiment. Each subject's perception of the syllables was assessed behaviorally outside of the MRI scanner. The following trials were presented randomly intermixed: 10 trials of McGurk syllables (auditory "ba" + visual "ga"); 10 trials of congruent syllables (auditory "ba" + visual "ba" or auditory "da" + visual "da"); 10 auditory-only "ba"; and 10 auditory-only "da" syllables. Auditory stimuli were delivered through headphones at ~70 dB, and visual stimuli were presented on a computer screen. Subjects were instructed to watch the mouth movements and listen to the speaker. To ensure appropriate understanding of the instructions before the start of the syllable experiment, subjects were first presented with three sentences (i.e., "Sally ate the ice cream"), three single-syllable words (i.e., "drink"), and six syllables ("fa," "ja," "na," "ba," "ha," and "la"), and were asked to repeat the presented stimuli.

To assess perception, subjects were asked to repeat aloud the perceived syllable, and all spoken responses were recorded using a microphone. No constraints were 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 (Olson et al., 2002; Colin et al., 2005) and is more informative with respect to possible interindividual 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 (McGurk and MacDonald, 1976). Responses corresponding to "ba," the auditory stimulus, indicated that subjects did not perceive the McGurk effect.

Table 1. McGurk susceptibility and demographics for each subject

Subject	McGurk susceptibility	Gender	Age (years)
1	0%	M	12
2	0%	F	9
3	0%	F	6
4	10%	M	7
5	15%	M	10
6	40%	M	7
7	50%	F	12
8	80%	F	9
9	90%	F	6
10	90%	F	10
11	90%	F	11
12	90%	F	11
13	95%	M	9
14	100%	F	6
15	100%	M	11
16	100%	M	12
17	100%	F	7

F, Female; M, male.

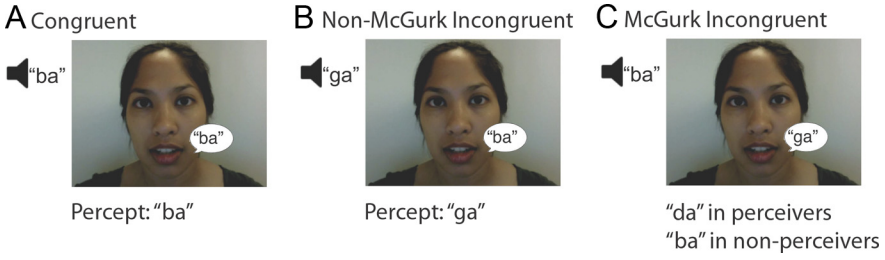

Figure 1. Audiovisual stimuli. ***A***, Congruent audiovisual syllable, consisting of matching auditory "ba" (depicted by speaker icon) and visual "ba" (single frame of video shown). Percept (shown below picture) is "ba." ***B***, Non-McGurk incongruent syllable, consisting of auditory "ga" and visual "ba." This stimulus does not result in an illusory percept; the resulting percept is most often "ga." ***C***, McGurk incongruent syllable, consisting of auditory "ba" and visual "ga." For McGurk perceivers, this results in the percept of an illusory "da." For nonperceivers, the percept is "ba."

All subjects participated in the behavioral experiment. Each subject was tested separately to reduce the possibility that the percept of other subjects could influence perception. Ten of these subjects underwent this testing on 2 separate days, 1 day for initial behavioral testing and another day for repeating the behavioral testing just before the fMRI experiment.

A cluster analysis on the behavioral data was performed to determine whether a subject fell into the perceiver or nonperceiver category. Each subject's percentage of McGurk responses was entered into the clusterdata function of MATLAB with the maximum number of categories to keep in the hierarchical tree set to two.

fMRI syllables experiment. Each fMRI scan series lasted for 4 min, and either two or three scan series were collected from each subject. Within each scan series, single syllables were presented within 2 s trials using a rapid event-related design. Each trial contained a video with a duration of 1.7–1.8 s, with fixation crosshairs occupying the remainder of the trial. Each scan series contained 25 McGurk trials, 25 incongruent trials, 25 congruent "ba" trials, 20 target trials (audiovisual "ma"), and 25 trials of fixation baseline. During fixation, the crosshairs were presented at the same position as the mouth during visual speech to minimize eye movements. Subjects were instructed to press a button during each target trial. Subjects identified target syllables with high precision (94% accuracy), indicating appropriate attention to the stimuli.

fMRI data analysis strategy. There is a debate in the literature on the merits of two different analysis strategies for fMRI data. In the first, termed the "SPM approach," a voxelwise whole-brain analysis is conducted to identify brain regions involved in the task of interest (Friston et al., 2006). In the second, termed the "functional localizer approach," independent functional localizers are used to identify one (or a few)

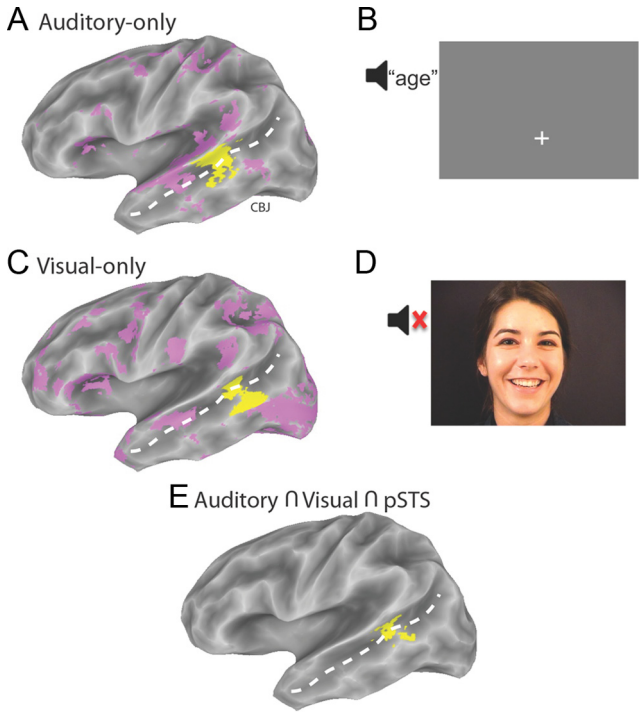

Figure 2. Responses to auditory and visual localizer stimuli in a single subject. ***A***, Cortical surface model from a single subject showing regions responding to blocks of auditory-only speech. Active regions within the posterior STS are shown in yellow, other active regions are shown in purple. Dashed white line indicates fundus of the STS. Three-letter code indicates anonymous subject identification ("CBJ"). ***B***, Blocks of auditory-only speech consisted of single words (represented with loudspeaker icon) and visual fixation crosshairs. ***C***, Cortical surface model from the same subject showing regions responding to blocks of visual-only (unisensory) faces. ***D***, Blocks of visual-only faces consisted of silent videos of emotional faces, illustrated by a single frame from a video. ***E***, Cortical surface model from the same subject showing regions within the posterior STS (pSTS) active during both auditory-only speech and visual-only faces.

functional brain regions (Saxe et al., 2006). The data from these independent functional localizers are then examined under the task of interest. While a debate on the relative merit of these approaches is beyond the scope of this manuscript (Friston et al., 2006; Saxe et al., 2006), they are not mutually exclusive. Therefore, we used both approaches to search for converging evidence on brain regions important for interindividual differences in audiovisual speech perception in children.

Whole-brain group analysis (SPM approach). The whole-brain group analysis was performed by normalizing each subject's average anatomical dataset to the 7–11-year-old child template from the NIH Pediatric MRI Data Repository (Fonov et al., 2011) using the auto_tlrc function in Analysis of Functional NeuroImages (AFNI) software. The output of the first-level individual subject regression analysis (t statistic of response to McGurk syllables) was smoothed using a $6 \times 6 \times 6$ mm FWHM Gaussian kernel and entered into the second-level multiple linear regression. The AFNI function 3dRegAna was used to identify voxels with a significant correlation between the response to McGurk stimuli and McGurk susceptibility across subjects.

Functional localizer analysis. For the functional localizer analysis, regions of interest (ROIs) were created from data collected in functional localizer scan series that were completely independent from the McGurk syllables scan series (Kriegeskorte et al., 2009; Vul et al., 2009). The functional localizer scan series contained eight blocks (four unisensory auditory and four unisensory visual in random order) with a duration of 20 s, with 10 s of fixation baseline between each block (Fig. 2). Each auditory block contained 10 2 s trials, one word per trial. Each visual block contained six 2.5 s trials, one emotional face per trial. Each block contained two target trials: the auditory target stimulus consisted of an auditory utterance of the word "press," while the visual target stimulus consisted of a visual smile. Subjects were instructed to press a button during each target trial. Subjects identified target syllables with high precision (92% accuracy for auditory and 94% accuracy for visual targets), indicating attention to the stimuli.

All ROIs were created on each subject's cortical surface based on activity during that subject's functional localizer scan, preventing any possible mismatch between brain and reference template (Yoon et al., 2009). The STS ROI was defined using a conjunction analysis to find all voxels that responded to both auditory words and visual faces significantly greater than baseline in the anatomically defined posterior STS ($T > 2$ for each modality) (Beauchamp, 2005; Beauchamp et al., 2008). The auditory cortex ROI was defined using the contrast of auditory words versus baseline ($T > 2$) to find active voxels within Heschl's gyrus (Patterson and Johnsrude, 2008; Upadhyay et al., 2008). The extrastriate visual cortex ROI was defined using the contrast of visual faces versus baseline ($T > 2$) within extrastriate lateral occipitotemporal cortex (Dumoulin et al., 2000). The fusiform face area ROI was created using the contrast of visual faces versus baseline within the FreeSurfer automated parcellation of fusiform gyrus and the adjacent lateral occipitotemporal sulcus (Kanwisher and Yovel, 2006).

Details of MRI and fMRI analysis. At the beginning of each scanning session, two T1-weighted MP-RAGE anatomical MRI scans were collected at 3 tesla using an 8-channel head gradient coil; the anatomical scans were aligned to each other and averaged to provide maximum gray–white contrast. Then, a cortical surface model was created with FreeSurfer (Dale et al., 1999; Fischl et al., 1999) to allow visualization and region-of-interest creation with SUMA (Argall et al., 2006). T2*-weighted images for fMRI were collected using gradient-echo echo-planar imaging (TR = 2015 ms, TE = 30 ms, flip angle = 90°) with in-plane resolution of 2.75 × 2.75 mm. Thirty-three 3 mm axial slices were collected, resulting in whole-brain coverage in most subjects. Each functional scan series consisted of 123 brain volumes. The first three volumes, collected before equilibrium magnetization was reached, were discarded resulting in 120 usable volumes. MRI-compatible in-ear headphones (Sensimetrics) covered with ear muffs were used to present auditory stimuli within the scanner. Visual stimuli were projected onto a screen using an LCD projector and viewed through a mirror attached to the head coil. Behavioral responses were collected using a fiber-optic button response pad (Current Designs). MR-compatible eye tracking (Applied Science Laboratories) was used in all fMRI experiments to ensure alertness and visual fixation.

fMRI data analysis was performed using AFNI software (Cox, 1996). Corrections for voxelwise multiple comparisons were performed using the false discovery rate procedure (Genovese et al., 2002) and reported as "q" values. Data were analyzed in each subject and then combined across subjects using a random-effects model. Functional data were aligned to the average anatomical dataset and motion corrected for each voxel in each subject using a local Pearson correlation (Saad et al., 2009). All analysis was performed in all voxels in each subject in the context of the generalized linear model using a maximum-likelihood approach using the AFNI function 3dDeconvolve. Movement covariates and baseline drifts (as second-order polynomials, one per scan series) were modeled as regressors of no interest. Head motion was quantified for each subject using two different measures of subject: average distance from mean position and peak deviation from mean position. Deconvolution with tent functions was used to separately estimate the complete hemodynamic response function (HRF) to each stimulus type in each voxel using nine tent functions that spanned the time between stimulus onset and 16 s after stimulus onset. As an additional control, the principle eigen time series of responses during McGurk stimuli was also extracted from each subject's STS ROI using the AFNI function 3dmaskSVD. For the ROI analysis, the average raw time series was created from all voxels in each ROI and then deconvolved with tent functions to extract the hemodynamic response from each ROI.

Correlations between fMRI data and behavioral measures. Simple Pearson correlation coefficients were calculated between the amplitude of the BOLD response (measured as the percentage signal change) and McGurk susceptibility (measured behaviorally). To ensure that correlations were not due to the effects of outliers, logistic-weighted regressions were calculated using the robustfit function in MATLAB. Correlations were also calculated between McGurk susceptibility and fMRI functional connectivity during McGurk perception. The amplitude of the hemodynamic response was estimated for each individual presentation of the McGurk stimulus and

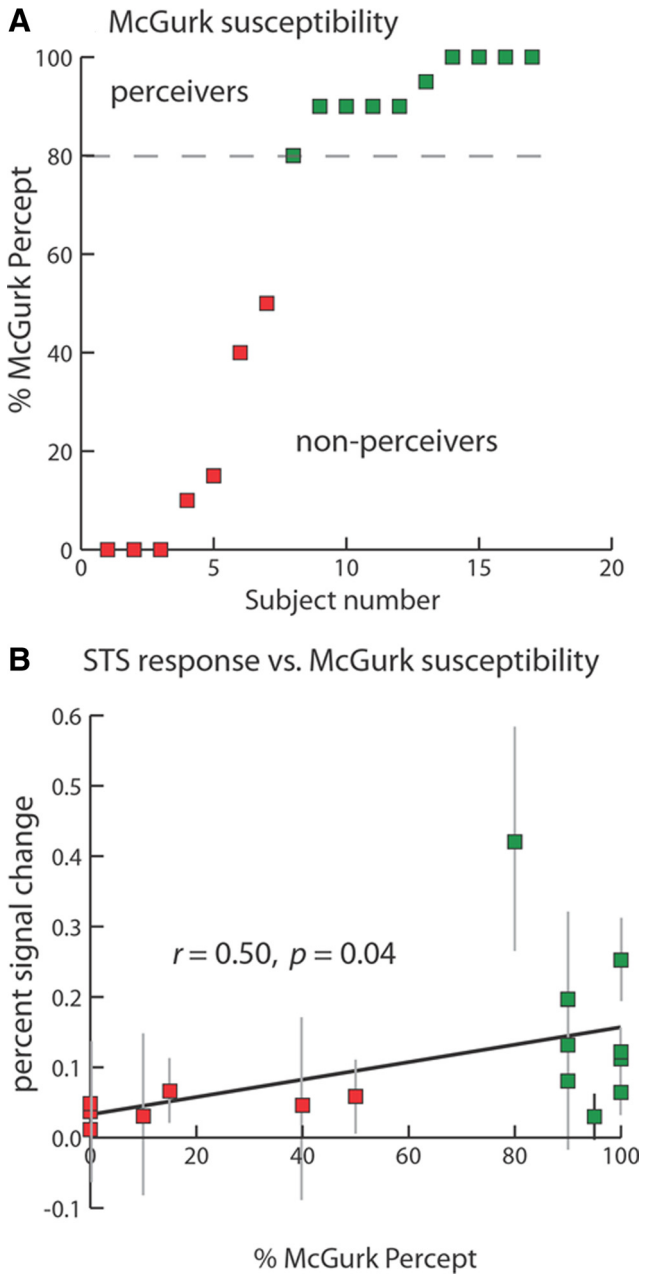

Figure 3. McGurk susceptibility and STS responses during McGurk syllables. ***A***, Behavioral data from all subjects, showing percentage of responses corresponding to the McGurk percept in each of 17 subjects, sorted by susceptibility. Dashed line indicates division into nonperceivers (red squares, <50% susceptibility) and perceivers (green squares, >80% susceptibility). ***B***, The BOLD fMRI response to McGurk syllables in the STS in each subject (gray bars show the SEM across trials) is plotted against that subject's McGurk susceptibility. There was a significant positive correlation between the STS response and the likelihood of experiencing the McGurk percept ($r = 0.50, p = 0.04$).

averaged within each ROI to produce a vector of 50 amplitudes, one per stimulus, as described in a previous study (Nath and Beauchamp, 2011). These amplitudes were used to estimate functional connection weights between areas (McIntosh and Gonzalez-Lima, 1994; Horwitz, 2003).

Results

Behavioral experiment

An open-choice behavioral experiment was conducted to measure each child's perception of the McGurk stimuli. There was a high degree of interindividual variability in McGurk susceptibility (Fig. 3*A*). Subjects 1–3 never reported the McGurk percept (0% McGurk percept), while subjects 14–17 always reported it (100% McGurk percept). Subjects were classified into two groups based on a cluster analysis of the McGurk percepts: nonperceivers (7 subjects, susceptibility 0–50%, mean value 16%); and strong perceivers (10 subjects, susceptibility 51–100%, mean value 94%). To determine whether McGurk susceptibility was stable ("state or trait"), we repeated testing on separate days for 10 subjects. There was not a significant difference in McGurk susceptibility on the different testing days (mean difference 9%, $p = 0.13$ using a paired t test) and no difference in group assignment for any subject, suggesting that McGurk susceptibility is stable within each individual subject. Although the perception of McGurk stimuli was very different between perceivers and nonperceivers, perception of nonillusory stimuli was similar, with high accuracy in both groups (mean 89% correct for audiovisual congruent syllables and 75% correct for auditory-only syllables).

Table 2. Whole-brain group analysis of correlation between BOLD response to McGurk stimuli and McGurk susceptibility

		Talairach coordinates (mm)		
	Size (mm^3)	*x*	*y*	*z*
Regions with positive correlation				
L STS	1816	−49	−45	5
R fusiform gyrus (FFA)	1566	13	−93	−9
L fusiform gyrus (FFA)	499	−43	−63	−23
Regions with negative correlation				
R IFG	72,322	47	25	17
R STG (auditory association areas)	12,757	51	−11	−15
L STG (auditory association areas)	9375	−41	−21	−1
L ventral occipitotemporal cortex	7900	−23	−83	−19
L post-central gyrus	1476	−33	−29	49
R ventral occipitotemporal cortex	1385	32	−62	−20
R middle occipital gyrus	1203	33	−81	7

R, Right; L, left; IFG, interior frontal gyrus; STG, superior temporal gyrus.

Whole-brain group analysis

We performed a voxelwise, whole-brain group analysis in which subjects' McGurk susceptibility scores were correlated with the BOLD response in each voxel in standard space to the presentation of three types of audiovisual speech: McGurk syllables, non-McGurk incongruent syllables that do not produce a McGurk effect, and congruent audiovisual syllables. Across stimulus types, the only positive correlation observed was with the response to McGurk syllables. The BOLD response to McGurk syllables in the left STS and the left and right fusiform gyri increased with McGurk susceptibility scores (Table 2; Fig. 4). No such relationship was observed for the other stimulus types.

Functional localizer analysis

In parallel with the whole-brain group analysis, we used an independent functional localizer scan to create eight ROIs, including left and right STS, auditory cortex, fusiform gyrus, and extrastriate visual cortex (Table 3, coordinates and volumes). Our initial analysis focused on the amplitude of the BOLD response to the three audiovisual syllable types in the left STS. The average STS response was 0.11% ± 0.03% to McGurk syllables, 0.08% ± 0.01% to incongruent syllables, and 0.13% ± 0.05% to congruent syllables. To examine the differences in activity for the three stimulus types in McGurk perceivers and nonperceivers, we performed a two-way ANOVA with the STS response as our dependent measure. The first factor was the McGurk susceptibility group determined from behavioral testing (strong perceivers and nonperceivers). The second factor was the stimulus condition (McGurk, incongruent,

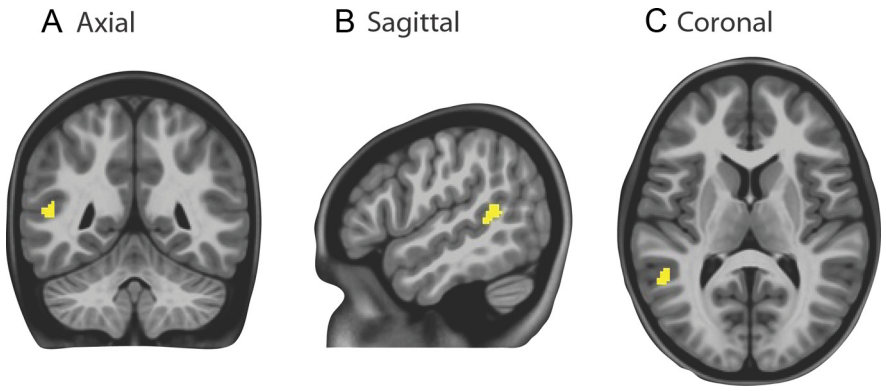

Figure 4. Whole-brain regression of fMRI responses and McGurk susceptibility. ***A***, The BOLD response to McGurk stimuli in each voxel in standard space was regressed against subjects' McGurk susceptibility. The largest cluster of voxels (shown in yellow) showing a positive correlation between BOLD response and behavior across subjects was located in the left STS. Anatomical underlay is the 7–11-year-old child template from the NIH Pediatric MRI Data Repository used for anatomical normalization (Fonov et al., 2011). ***B***, Left STS voxels (in yellow) showing a positive correlation between BOLD response and behavior, shown on a sagittal slice. ***C***, Coronal slice showing STS voxels with a positive BOLD– behavior correlation.

Table 3. Regions of interest created from auditory and visual localizers

		Talairach coordinates (mm)		
	Size (mm^3)	*x*	*y*	*z*
Auditory localizer				
L STG	1335 ± 203	−45 ± 1	−3 ± 2	26 ± 4
R STG	973 ± 147	46 ± 1	−2 ± 2	24 ± 4
Visual localizer				
L fusiform gyrus	1659 ± 282	−36 ± 1	−42 ± 2	7 ± 4
R fusiform gyrus	1979 ± 371	33 ± 1	−40 ± 2	6 ± 4
L extrastriate	1154 ± 143	−42 ± 2	−47 ± 3	21 ± 4
R extrastriate	1248 ± 173	39 ± 1	−48 ± 1	20 ± 5
Auditory and visual				
L STS	413 ± 83	−51 ± 1	−30 ± 2	28 ± 5
R STS	750 ± 155	47 ± 1	−25 ± 2	28 ± 4

STG, Superior temporal gyrus. ROIs were created separately in each subject. Values represent mean ± SEM.

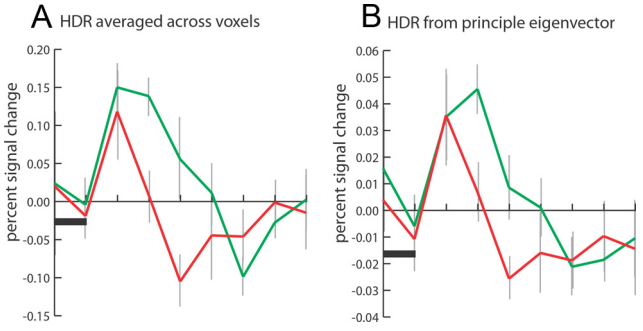

Figure 5. Time course of STS response. ***A***, The plot shows the time course of the STS response to a single presentation of a McGurk syllable, extracted using a finite impulse response deconvolution procedure. The *y*-axis shows the BOLD percentage signal change. The *x*-axis is the time following stimulus onset (2 s per tick mark), and the black bar represents the 2 s duration of the McGurk syllable. The red line shows the STS response averaged across all nonperceivers. The green line shows the STS response averaged across all perceivers. The gray bars show the SEM at each time point. To generate these curves, in each subject the mean time series from all voxels in the STS ROI was calculated, followed by deconvolution. ***B***, The time course of the STS response to a McGurk syllable calculated using a different technique. In each subject, the principle eigen time series from all voxels in the STS ROI was calculated using singular value decomposition, followed by deconvolution.

and congruent syllables). The ANOVA showed a significant main effect of the McGurk susceptibility group ($F_{(1,45)} = 12.1$, $p = 0.001$), but not the stimulus condition ($F_{(2,45)} = 0.18$, $p = 0.84$), on the STS response. The main effect of group was driven by a larger response to McGurk syllables in the perceivers compared with the nonperceivers (0.16% ± 0.01 for perceivers vs 0.04% ± 0.01 for nonperceivers, $p = 0.02$). There was also a trend toward a larger response in perceivers for the other stimulus types (0.20% ± 0.07 for perceivers vs 0.03% ± 0.03 for nonperceivers, $p = 0.07$ for incongruent syllables; and 0.13% ± 0.03 vs 0.05% ± 0.03, $p = 0.11$ for congruent syllables).

Next, we examined individual STS responses to the stimuli. Across all subjects, there was a significant positive correlation between each subject's STS response to McGurk speech and their likelihood of experiencing the McGurk percept ($r = 0.50$, $p = 0.04$ with simple correlation; $p = 0.02$ with robust regression) (Fig. 3*B*). There was no correlation between STS response and McGurk susceptibility for non-McGurk incongruent ($r = 0.36$, $p = 0.16$; $p = 0.17$ with robust regression) or congruent syllables ($r = 0.38$, $p = 0.13$; $p = 0.24$ with robust regression).

In additional analyses, we examined the response to McGurk stimuli in seven additional ROIs (right STS, left and right fusiform gyri, left and right auditory cortex, and left and right extrastriate visual cortex) created from the independent localizer scans. Two ROIs showed the main effect of the susceptibility group: the left extrastriate visual cortex ($F_{(1,45)} = 8.1$, $p = 0.007$) and the left fusiform gyrus ($F_{(1,45)} = 6.8$, $p = 0.01$) were driven by larger responses in perceivers. The other ROIs showed no main effects of McGurk susceptibility group, and no areas exhibited a main effect of stimulus condition or an interaction between McGurk susceptibility group or stimulus condition.

Possible differences in functional connectivity

To better understand the activity differences between perceivers and nonperceivers, we performed a functional connectivity analysis. A two-way ANOVA on the connection weights (with areas and McGurk susceptibility group as factors) revealed a significant main effect of ROI ($F_{(4,75)} = 5.6$, $p = 0.0006$) but not of McGurk susceptibility group ($F_{(1,75)} = 0.6$, $p = 0.45$). The main effect of ROI was driven by strong connections between STS and left MT ($r = 0.61$), STS and auditory cortex ($r = 0.57$), STS and left fusiform gyrus ($r = 0.46$), and left and right fusiform gyrus ($r = 0.67$), compared with the weak connection between left STS and right fusiform gyrus ($r = 0.31$).

Differences in hemodynamic response function between groups

In our initial analysis, we measured the response to each stimulus using the standard technique of fitting each response with a canonical reference gamma function waveform and extracting a single estimate of the response to each stimulus in each voxel. To extract more information about the shape of the response in the left STS, we performed an additional analysis in which we separately fit impulse response functions (also known as delta or tent functions) to measure the complete time course of the response to each stimulus.

The hemodynamic responses to McGurk stimuli are shown in perceivers and nonperceivers in Figure 5*A*. In addition to the amplitude difference detected by our initial analysis, there were

striking differences in the appearance of the HRF. The HRF in nonperceivers peaked at 4 s and decayed to baseline at 6 s, followed by a substantial negative deflection (undershoot) below baseline consistent with previous studies (Buxton et al., 1998; Richter and Richter, 2003; Schroeter et al., 2006). In contrast, the HRF in perceivers peaked at 4 s, remaining elevated for an extended duration until returning to baseline at 10 s poststimulus (followed by an undershoot similar to that of nonperceivers). To quantify this difference, we calculated the integral of the hemodynamic response function by summing the response in a window between 2 and 10 s. This integral was significantly greater in the perceivers than in the nonperceivers (0.35% ± 0.10% vs −0.04% ± 0.09%, $p = 0.01$).

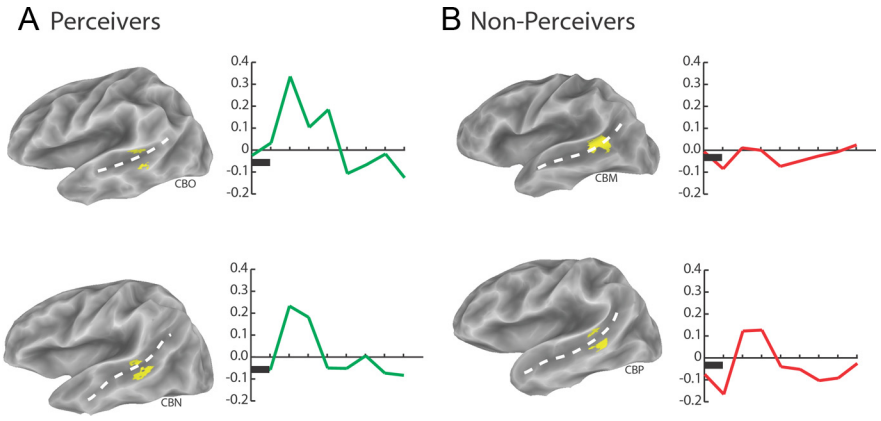

Figure 6. STS responses in individual perceivers and nonperceivers. ***A***, STS responses in two individual McGurk perceivers (three-letter code indicates anonymous subject identification). Active regions in the STS responding to unisensory auditory and visual stimuli shown is in yellow (see Fig. 2). Time course of STS response to single McGurk syllables is shown in green (see Fig. 5). Same *y*-axis scale (percentage BOLD signal change) in all plots. ***B***, STS responses in two individual nonperceivers. Time course of STS response to single McGurk syllables is shown in red.

We next compared the time course of activity at each time point in the window using an ANOVA with group (perceivers vs nonperceivers) and poststimulus time as factors. The ANOVA revealed significant main effects of group ($F_{(1,75)} = 8.0, p = 0.006$) and poststimulus time ($F_{(4,75)} = 5.1, p = 0.001$) without a significant interaction. *Post hoc t* tests revealed that the amplitude of response than in non-McGurk perceivers was greater at 6 and 8 s poststimulus (0.14% ± 0.03% vs 0.01% ± 0.04%, $p = 0.01$ at 6 s; 0.06% ± 0.06% vs − 0.10% ± 0.04%, $p = 0.04$ at 8 s), demonstrating a significantly prolonged hemodynamic response to McGurk stimuli in perceivers.

Our initial analysis used the mean time series across all voxels in the STS ROI. However, if the ROI is heterogeneous, the eigen time series can give a better representation of the hemodynamic response. Therefore, we repeated the analysis, extracting the eigen time series from the STS instead of the mean time series. A similar difference in the time course of activity between perceivers and nonperceivers was observed (Fig. 5*B*). The integral between 2 and 10 s was significantly greater in perceivers than in nonperceivers (0.08% ± 0.03% vs −0.01% ± 0.02%, $p = 0.03$), with the ANOVA showing significant main effects of group ($F_{(1,75)} = 5.2, p = 0.03$) and poststimulus time ($F_{(4,75)} = 5.5, p = 0.0006$) without a significant interaction. *Post hoc t* tests found that the amplitude of response was greater in perceivers at 6 and 8 s poststimulus (0.05% ± 0.01% vs 0.007% ± 0.01%, $p = 0.02$ at 6 s; 0.01% ± 0.01% vs −0.03% ± 0.008%, $p = 0.04$ at 8 s).

Possible age effects

Because our subjects ranged in age from 6 to 12 years, a time period in which there are large developmental changes, we examined our data for age effects. Two significant effects of age were noted. The amount of head motion was negatively correlated with age, with older subjects exhibiting less head motion than younger subjects (mean head motion = 0.82 mm, correlated with age in years: $r = -0.50$, $p = 0.04$; average maximum head motion = 2.7 mm, correlation with age: $r = -0.56$, $p = 0.02$). Performance on the in-scanner task during the localizer runs was positively correlated with age, with older subjects identifying the target words and faces more accurately (mean percentage correct = 94%, correlation with age: $r = 0.49$, $p = 0.046$). We found no significant age effects in the remainder of the data. There was no correlation found between McGurk susceptibility and age ($r = 0.08$, $p = 0.76$). There was no correlation between response amplitude and age for any stimulus category or ROI, and neither the size nor the laterality of the STS ROIs varied with age.

Possible gender effects

We examined our data for any differences in behavioral responses or fMRI activity between males and females using unpaired *t* tests. There was no gender difference between males and females in McGurk susceptibility, in-scanner task accuracy, handedness score, mean or maximum head motion, or size or amplitude of response in the STS ROIs.

Other possible confounds

These data suggested a link between activity in the STS and susceptibility to the McGurk effect. We wished to rule out other possible confounds that could lead to differences between groups, such as head movements, or nonspecific differences in BOLD amplitude (perhaps due to attention or arousal). We found no correlation between STS response and mean or peak head movements, and no correlation between the STS response to McGurk stimuli and the BOLD amplitude in localizer blocks (as would be expected if, for instance, nonperceivers simply never attended to the stimulus). There was no correlation between the STS response to McGurk stimuli and performance on the in-scanner task consisting of detection of "ma" syllables. There was no difference in the size of STS ROIs between perceivers and nonperceivers (362 vs 486 mm^3, $p = 0.48$) (Fig. 6).

Discussion

To understand the neural substrates of interindividual differences in children's perception of audiovisual speech, we conducted behavioral and fMRI experiments. All children accurately perceived auditory-only auditory syllables and congruent audiovisual syllables, but individual differences were observed in the perception of McGurk syllables: 59% of subjects were susceptible to the McGurk effect (perceiving it most of the time), while 41% were not. McGurk susceptibility was stable across testing sessions, suggesting that it reflects a difference in multisensory speech perception between individuals, rather than day-to-day variability within individual subjects. The interindividual variability we observed in McGurk perception is similar to that previously reported in both children (Schorr et al., 2005) and adults

(MacDonald et al., 2000; Gentilucci and Cattaneo, 2005; Benoit et al., 2010; Schwartz, 2010).

Converging evidence from two very different fMRI analysis approaches—whole-brain analysis and functional localizers—pointed to the STS as an important brain area underlying inter-individual differences in audiovisual speech perception. Both approaches found that the left STS was more active in children that perceived the McGurk effect than in those that did not. A simple explanation for this is that increased activity in the STS is a neural signature for multisensory integration. Perception of the McGurk effect requires that the incongruent auditory and visual components of the stimulus be integrated to form a new percept that is compatible with both the auditory and visual stimuli. Perceivers integrate the auditory and visual components of the McGurk stimulus, resulting in STS activity. Nonperceivers do not integrate the modalities and display little or no STS activity, resulting in a percept that is most commonly the auditory component of the stimulus ("ba") (McGurk and MacDonald, 1976). A causal relationship between STS activity and McGurk perception was demonstrated in a recent TMS study: when the STS was disrupted, McGurk-perceiving adults became more similar to nonperceivers, reporting an auditory percept instead of a McGurk percept (Beauchamp et al., 2010). Converging evidence is also provided by a recent fMRI study of McGurk perception in adults (Nath and Beauchamp, 2011), in which greater left STS activity was observed in perceivers compared with nonperceivers, just as was found in children in the present study. In adults, a positive correlation between STS activity and McGurk susceptibility was observed ($r = 0.73$, $p = 0.003$). The same positive correlation was observed in children ($r = 0.50$, $p = 0.04$) but with a lower slope, due to the fact that some child perceivers had weak STS responses (minimum value = 0.03%), while all adult perceivers had strong STS responses (minimum value = 0.17%). This result is paradoxical: if the STS is critical for audiovisual speech integration and McGurk susceptibility, how can McGurk perception occur in the absence of strong STS activity in some child subjects?

A possible explanation can be found by examining other areas outside the STS. In children, activity in the left fusiform gyrus and the left extrastriate visual cortex was significantly greater in perceivers than nonperceivers, while activity in auditory cortex was negatively correlated with McGurk susceptibility. The increased fusiform and extrastriate visual activity in children is intriguing. Both of these ROIs showed stronger activity in child McGurk perceivers, and they are part of an extended network for moving faces linked to lip-reading ability in adults (Calvert and Campbell, 2003; Ruytjens et al., 2006). An important contributor to the McGurk effect may be the ability to speech read (decode speech from visual cues); therefore, increased activity in these areas may reflect greater visual processing of mouth movements in McGurk perceivers. In contrast, activity in auditory areas was negatively correlated with McGurk susceptibility. If some children perform more auditory processing of audiovisual speech (as indicated by greater amplitude of response in auditory cortex) and less visual processing (as indicated by weaker amplitude of response in visual cortex), they would not be expected to perceive the McGurk syllable. Indeed, the most common percept reported by nonperceivers is the auditory syllable.

In adults, a very different pattern was observed. There was no difference in adult fusiform activity between perceivers and nonperceivers, and activity in extrastriate visual cortex was negatively correlated with McGurk susceptibility while activity in auditory cortex was positively correlated with susceptibility (the reverse of the effect seen in children). This suggests very different functional interactions between speech-processing areas in adults and children, as observed in previous studies of language networks (Dick et al., 2010; Zielinski et al., 2010). Animal studies have demonstrated that multisensory areas are not mature until relatively late in development (Wallace et al., 2006), consistent with other studies showing differences between child and adult language networks for the same age ranges that we examined in our study (Schlagger et al., 2002; Turkeltaub et al., 2003; Booth et al., 2004; Brown et al., 2005, 2006; Dick et al., 2010). The STS and other regions of the language network, including Broca's area, mature as a functional unit in 1–4-month-old infants (Leroy et al., 2011), supporting the idea of a distributed network for language processing, even in very young children. Our finding of a link between brain activity (the STS) and language behavior (McGurk perception) is also compatible with other studies that have compared neural activity and language abilities (Shaywitz et al., 2002; Turkeltaub et al., 2004; Schlagger and McCandliss, 2007; Hoeft et al., 2011).

In addition to the finding of greater amplitude of activity within the left STS in McGurk perceivers, we also found a difference in the shape of the hemodynamic response between children who perceived the McGurk effect and those who did not. The HRF of perceivers was significantly extended in time relative to that of nonperceivers. In our studies of adults, we did not observe this effect. The broader HRFs of child McGurk perceivers versus adult McGurk perceivers is consistent with prior studies that have found a broader hemodynamic response in children than in adults (Richter and Richter, 2003). In general, prolonged neural responses in children may reflect neural plasticity, as synaptic weights are strengthened, and changes in neurovascular coupling (Harris et al., 2011). In the perceivers, this could reflect learning of the association between auditory phonemes and visual visemes within the STS.

We found no consistent effects of subject age on McGurk perception, either between children and adults, or between children of different ages. In our study, 59% of children perceived the McGurk effect, similar to the 57% of child perceivers in a study of 5–14-year-olds (Schorr et al., 2005) and the 50% of adult perceivers we observed in a study using identical stimuli and testing protocols ($p = 0.31$ using the binomial distribution). Some studies suggest that McGurk susceptibility increases with age (McGurk and MacDonald, 1976; Hockley and Polka, 1994; Dupont et al., 2005; Tremblay et al., 2007). One possible explanation for the lack of age effects in the present study is a small sample size [31 total children and adult subjects in our studies vs 103 total subjects in the original study by McGurk and MacDonald (1976)].

While multisensory speech comprises only a small subset of language, there is evidence of a link between changes in audiovisual speech perception and language development. Infants are able to acquire sufficient auditory experience with speech *in utero* to identify the prosodic patterns of their native language (Mehler et al., 1978, 1988; Moon et al., 1993; Nazzi et al., 1998; Ramus et al., 2000). As early as 4 months of age, infants are able to differentiate visual speech of their own native language from a non-native exemplar (Weikum et al., 2007). At this same early age, infants acquire the ability to synthesize auditory and visual speech into one percept (Lewkowicz, 2000). For example, infants at 2 months of age match lip and voice synchrony (Dodd, 1979), and infants at 4.5 months of age are capable of matching the auditory and visual attributes of speech syllables (Kuhl and Meltzoff, 1982; Patterson and Werker, 1999). Audiovisual speech perception is critical to many aspects of perceptual, cognitive, and

social learning (Rochat, 1999; Gibson and Pick, 2000) and requires many years of experience and feedback. The infant behavioral literature carefully documents the gradual development of audiovisual speech perception in terms of numerous factors such as spectral information, temporal synchrony, affect, gender, and age of speaker (Bahrick et al., 2005). Thus, the examination of multisensory speech perception may reveal important clues about the development of speech perception and language defined more broadly.

We are not aware of any longitudinal studies of McGurk perception. While we observed stable McGurk perception across a relatively short interval (approximately weeks), it is fascinating to speculate whether over a longer time span any of our nonperceivers could "convert" to perceivers, either through natural development or with a training program that focused on speech reading (Gagné et al., 1991; Bernstein et al., 2001; Blumsack et al., 2007).

References

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

Bahrick LE, Hernandez-Reif M, Flom R (2005) The development of infant learning about specific face–voice relations. Dev Psychol 41:541–552.

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

Beauchamp MS, Lee KE, Argall BD, Martin A (2004) Integration of auditory and visual information about objects in superior temporal sulcus. Neuron 41:809–823.

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

Beauchamp MS, Nath AR, 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.

Belin P, Zatorre RJ, Ahad P (2002) Human temporal-lobe response to vocal sounds. Brain Res Cogn Brain Res 13:17–26.

Benoit MM, Raij T, Lin FH, Jääskeläinen IP, Stufflebeam S (2010) Primary and multisensory cortical activity is correlated with audiovisual percepts. Hum Brain Mapp 31:526–538.

Bernstein LE, Auer ET Jr, Tucker PE (2001) Enhanced speechreading in deaf adults: can short-term training/practice close the gap for hearing adults? J Speech Lang Hear Res 44:5–18.

Blumsack JT, Bower CR, Ross ME (2007) Comparison of speechreading training regimens. Percept Mot Skills 105:988–996.

Booth JR, Burman DD, Meyer JR, Gitelman DR, Parrish TB, Mesulam MM (2004) Development of brain mechanisms for processing orthographic and phonologic representations. J Cogn Neurosci 16:1234–1249.

Brown TT, Lugar HM, Coalson RS, Miezin FM, Petersen SE, Schlaggar BL (2005) Developmental changes in human cerebral functional organization for word generation. Cereb Cortex 15:275–290.

Brown TT, Petersen SE, Schlaggar BL (2006) Does human functional brain organization shift from diffuse to fodal with development? Dev Sci 9:9–11.

Burnham D, Dodd B (2004) Auditory-visual speech integration by prelinguistic infants: perception of an emergent consonant in the McGurk effect. Dev Psychobiol 45:204–220.

Buxton RB, Wong EC, Frank LR (1998) Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. Magn Reson Med 39:855–864.

Callan DE, Jones JA, Munhall K, Kroos C, Callan AM, Vatikiotis-Bateson E (2004) Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. J Cogn Neurosci 16:805–816.

Calvert GA, Campbell R (2003) Reading speech from still and moving faces: the neural substrates of visible speech. J Cogn Neurosci 15:57–70.

Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr Biol 10:649–657.

Cao F, Khalid K, Lee R, Brennan C, Yang Y, Li K, Bolger DJ, Booth JR (2010) Development of brain networks involved in spoken work processing of Mandarin Chinese. Neuroimage 57:750–759.

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.

Cone NE, Burman DD, Bitan T, Bolger DJ, Booth JR (2008) Developmental changes in brain regions involved in phonological and orthographic processing during spoken language processing. Neuroimage 41:623–635.

Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 29:162–173.

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

Devlin JT, Jamison HL, Gonnerman LM, Matthews PM (2006) The role of the posterior fusiform gyrus in reading. J Cogn Neurosci 18:911–922.

Dick AS, Solodkin A, Small SL (2010) Neural development of networks for audiovisual speech comprehension. Brain Lang 114:101–114.

Dodd B (1979) Lip reading in infants: attention to speech presented in- and out-of-synchrony. Cogn Psychol 11:478–484.

Dumoulin SO, Bittar RG, Kabani NJ, Baker CL Jr, Le Goualher G, Bruce Pike G, Evans AC (2000) A new anatomical landmark for reliable identification of human area V5/MT: a quantitative analysis of sulcal patterning. Cereb Cortex 10:454–463.

Dupont S, Aubin J, Menard L (2005) A study of the McGurk effect in 4 and 5-year-old French Canadian children. ZAS Papers Linguistics 40:1–17.

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

Fonov V, Evans AC, Botteron K, Almli CR, McKinstry RC, Collins DL (2011) Unbiased average age-appropriate atlases for pediatric studies. Neuroimage 54:313–327.

Friston KJ, Rotshtein P, Geng JJ, Sterzer P, Henson RN (2006) A critique of functional localisers. Neuroimage 30:1077–1087.

Gagné JP, Dinon D, Parsons J (1991) An evaluation of CAST: a Computer-Aided Speechreading Training program. J Speech Hear Res 34:213–221.

Genovese CR, Lazar NA, Nichols T (2002) Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage 15:870–878.

Gentilucci M, Cattaneo L (2005) Automatic audiovisual integration in speech perception. Exp Brain Res 167:66–75.

Gibson EJ, Pick A (2000) An ecological approach to perceptual learning and development. New York: Oxford UP.

Harris JJ, Reynell C, Attwell D (2011) The physiology of developmental changes in BOLD functional imaging signals. Dev Cogn Neurosci 1:199–216.

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

Hoeft F, McCandliss BD, Black JM, Gantman A, Zakerani N, Hulme C, Lyytinen H, Whitfield-Gabrieli S, Glover GH, Reiss AL, Gabrieli JD (2011) Neural systems predicting long-term outcome in dyslexia. Proc Natl Acad Sci U S A 108:361–366.

Horwitz B (2003) The elusive concept of brain connectivity. Neuroimage 19:466–470.

Ishai A, Schmidt CF, Boesiger P (2005) Face perception is mediated by a distributed cortical network. Brain Res Bull 67:87–93.

Kanwisher N, Yovel G (2006) The fusiform face area: a cortical region specialized for the perception of faces. Philos Trans R Soc Lond B Biol Sci 361:2109–2128.

Kessler H, Doyen-Waldecker C, Hofer C, Hoffmann H, Traue HC, Abler B (2011) Neural correlates of the perception of dynamic versus static facial expressions of emotion. Psychosoc Med 8:Doc03.

Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. Nat Neurosci 12:535–540.

Kuhl PK, Meltzoff AN (1982) The bimodal perception of speech in infancy. Science 218:1138–1141.

Leroy F, Glasel H, Dubois J, Hertz-Pannier L, Thirion B, Mangin JF, Dehaene-Lambertz G (2011) Early maturation of the linguistic dorsay pathway in human infants. J Neurosci 31:1500–1506.

Lewkowicz DJ (2000) The Development of intersensory temporal perception: an epigenetic systems/limitations view. Psychological Bulletin 126:281–308.

Lidzba K, Schwilling E, Grodd W, Krageloh-Mann I, Wilke M (2011) Language comprehension vs. language production: age effects on fMRI activation. Brain Lang. Advance online publication. Retrieved August 18, 2001. doi:10.1016/j.bandl.2011.02.003.

MacDonald J, Andersen S, Bachmann T (2000) Hearing by eye: how much spatial degradation can be tolerated? Perception 29:1155–1168.

MacSweeney M, Woll B, Campbell R, McGuire PK, David AS, Williams SC, Suckling J, Calvert GA, Brammer MJ (2002) Neural systems underlying British Sign Language and audio-visual English processing in native users. Brain 125:1583–1593.

Massaro DW, Thompson LA, Barron B, Laren E (1986) Developmental changes in visual and auditory contributions to speech perception. J Exp Child Psychol 41:91–113.

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

McIntosh AR, Gonzalez-Lima F (1994) Structural equation modeling and its application to network analysis in functional brain imaging. Hum Brain Mapp 2:2–22.

Mehler J, Bertoncini J, Barriere M (1978) Infant recognition of mother's voice. Perception 7:491–497.

Mehler J, Jusczyk P, Lambertz G, Halsted N, Bertoncini J, Amiel-Tison C (1988) A precursor of language acquisition in young infants. Cognition 29:143–178.

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

Moon C, Panneton-Cooper R, Fifer WP (1993) Two-day olds prefer their native language. Infant Behav Dev 16:495–500.

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

Nath AR, Beauchamp MS (2011) A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. Neuroimage. Advance online publication. Retrieved August 18, 2011. doi:10.1016/j.neuroimage.2011.07.024.

Nazzi T, Bertoncini J, Mehler J (1998) Language discrimination by newborns: toward an understanding of the role of rhythm. J Exp Psychol Hum Percept Perform 24:756–766.

Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9:97–113.

Olson IR, Gatenby JC, Gore JC (2002) A comparison of bound and unbound audio-visual information processing in human cerebral cortex. Cogn Brain Res 14:129–138.

Patterson ML, Werker JF (1999) Matching phonetic information in lips and voice is robust in 4.5-month-old infants. Infant Behav Dev 22:237–247.

Patterson RD, Johnsrude IS (2008) Functional imaging of the auditory processing applied to speech sounds. Philos Trans R Soc Lond B Biol Sci 363:1023–1035.

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

Ramus F, Hauser MD, Miller C, Morris D, Mehler J (2000) Language discrimination by human newborns and by cotton-top tamarin monkeys. Science 288:349–351.

Richter W, Richter M (2003) The shape of the fMRI BOLD response in children and adults changes systematically with age. Neuroimage 20:1122–1131.

Rochat P (1999) Early social cognition: understanding others in the first months of life. Mahwah, NJ: Erlbaum.

Rosenblum LD, Schmuckler MA, Johnson JA (1997) The McGurk effect in infants. Percept Psychophys 59:347–357.

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

Saad ZS, Glen DR, Chen G, Beauchamp MS, Desai R, Cox RW (2009) A new method for improving functional-to-structural MRI alignment using local Pearson correlation. Neuroimage 44:839–848.

Sabatinelli D, Fortune EE, Li Q, Siddiqui A, Krafft C, Oliver WT, Beck S, Jeffries J (2011) Emotional perception: meta-analyses of face and natural scene processing. Neuroimage 54:2524–2533.

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

Schlaggar BL, McCandliss BD (2007) Development of neural systems for reading. Annu Rev Neurosci 30:475–503.

Schlaggar BL, Brown TT, Lugar HM, Visscher KM, Miezin FM, Petersen SE (2002) Functional neuroanatomical differences between adults and school-age children in the processing of single words. Science 296:1476–1479.

Schorr EA, Fox NA, van Wassenhove V, Knudsen EI (2005) Audio-visual fusion in speech perception in children with cochlear implants. Proc Natl Acad Sci U S A 102:18748–18750.

Schroeter ML, Kupka T, Mildner T, Uludağ K, von Cramon DY (2006) Investigating the post-stimulus undershoot of the BOLD signal-a simultaneous fMRI and fNIRS study. Neuroimage 30:349–358.

Schwartz JL (2010) A reanalysis of McGurk data suggests that audiovisual fusion in speech perception is subject-dependent. J Acoust Soc Am 127:1584–1594.

Sekiyama K, Burnham D (2008) Impact of language on development of auditory-visual speech perception. Dev Sci 11:306–320.

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

Shaywitz BA, Shaywitz SE, Pugh KR, Mencl WE, Fulbright RK, Skudlarski P, Constable RT, Marchione KE, Fletcher JM, Lyon GR, Gore JC (2002) Disruption of posterior brain systems for reading in children with developmental dyslexia. Biol Psychiatry 52:101–110.

Stevenson RA, James TW (2009) Audiovisual integration in human superior temporal sulcus: inverse effectiveness and the neural processing of speech and object recognition. Neuroimage 44:1210–1223.

Tremblay C, Champoux F, Voss P, Bacon BA, Lepore F, Théoret H (2007) Speech and non-speech audio-visual illusions: a developmental study. PLoS ONE 2:e742.

Turkeltaub PE, Gareau L, Flowers DL, Zeffiro TA, Eden GF (2003) Development of neural mechanisms for reading. Nat Neurosci 6:767–773.

Turkeltaub PE, Flowers DL, Verbalis A, Miranda M, Gareau L, Eden GF (2004) The neural basis of hyperlexic reading: an fMRI case study. Neuron 41:11–25.

Upadhyay J, Silver A, Knaus TA, Lindgren KA, Ducros M, Kim DS, Tager-Flusberg H (2008) Effective and structural connectivity in the human auditory cortex. J Neurosci 28:3341–3349.

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 Linden S, Vroomen J (2008) Audiovisual speech recalibration in children. J Child Lang 35:809–822.

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.

Wallace MT, Carriere BN, Perrault TJ Jr, Vaughan JW, Stein BE (2006) The development of cortical multisensory integration. J Neurosci 26:11844–11849.

Weikum WM, Vouloumanos A, Navarra J, Soto-Faraco S, Sebastián-Gallés N, Werker JF (2007) Visual language discrimination in infancy. Science 316:1159.

Wilson M (1988) The MRC psycholinguistic database: machine readable dictionary, version 2. Behav Res Methods Instruments Comput 20:6–11.

Wright TM, Pelphrey KA, Allison T, McKeown MJ, McCarthy G (2003) Polysensory interactions along lateral temporal regions evoked by audiovisual speech. Cereb Cortex 13:1034–1043.

Yoon U, Fonov VS, Perusse D, Evans AC (2009) The effect of template choice on morphometric analysis of pediatric brain data. Neuroimage 45:769–777.

Zielinski BA, Gennatas ED, Zhou J, Seeley WW (2010) Network-level structural covariance in the developing brain. Proc Natl Acad Sci U S A 107:18191–18196.