Neural development of networks for audiovisual speech comprehension

Anthony Steven Dick [a,*], Ana Solodkin [a], Steven L. Small [a,b]

[a] Department of Neurology, The University of Chicago, IL, USA
[b] Department of Psychology, The University of Chicago, IL, USA

ABSTRACT

Everyday conversation is both an auditory and a visual phenomenon. While visual speech information enhances comprehension for the listener, evidence suggests that the ability to benefit from this information improves with development. A number of brain regions have been implicated in audiovisual speech comprehension, but the extent to which the neurobiological substrate in the child compares to the adult is unknown. In particular, developmental differences in the network for audiovisual speech comprehension could manifest through the incorporation of additional brain regions, or through different patterns of effective connectivity. In the present study we used functional magnetic resonance imaging and structural equation modeling (SEM) to characterize the developmental changes in network interactions for audiovisual speech comprehension. The brain response was recorded while children 8- to 11-years-old and adults passively listened to stories under audiovisual (AV) and auditory-only (A) conditions. Results showed that in children and adults, AV comprehension activated the same fronto-temporo-parietal network of regions known for their contribution to speech production and perception. However, the SEM network analysis revealed age-related differences in the functional interactions among these regions. In particular, the influence of the posterior inferior frontal gyrus/ventral premotor cortex on supramarginal gyrus differed across age groups during AV, but not A speech. This functional pathway might be important for relating motor and sensory information used by the listener to identify speech sounds. Further, its development might reflect changes in the mechanisms that relate visual speech information to articulatory speech representations through experience producing and perceiving speech.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

In naturalistic situations, such as conversation, the incoming auditory speech stream is accompanied by information from the face, particularly from the lips, mouth, and eyes of the speaker. This visual information has been shown to enhance speech comprehension in both children and adults (Binnie, Montgomery, & Jackson, 1974; Dodd, 1979; MacLeod & Summerfield, 1987; Massaro, 1984; Massaro, Thompson, Barron, & Laren, 1986; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007; Sumby & Pollack, 1954; Summerfield, 1979). Although sensitivity to visual speech information appears early in development (Burnham & Dodd, 2004; Kuhl & Meltzoff, 1982; Patterson & Werker, 2003; Rosenblum, Schmuckler, & Johnson, 1997; Teinonen, Aslin, Alku, & Csibra, 2008; Weikum et al., 2007), there is evidence that it continues to develop throughout childhood (Desjardins & Werker, 2004; Hockley & Polka, 1994; Massaro et al., 1986; McGurk & MacDonald, 1976; Sekiyama & Burnham, 2008; van Linden & Vroomen, 2008). With respect to

neurobiology, recent research suggests that audiovisual speech comprehension in adults is mediated by a neural network that incorporates primary sensory regions as well as posterior inferior frontal gyrus and ventral premotor cortex (IFGOp/PMv), supramarginal gyrus (SMG), posterior superior temporal gyrus (STGp), planum temporale (PTe), and the posterior superior temporal sulcus (STSp), and involves strong effective connectivity among these regions (Bernstein, Lu, & Jiang, 2008; Callan, Jones, Callan, & Akahane-Yamada, 2004; Callan et al., 2003; Calvert & Campbell, 2003; Calvert, Campbell, & Brammer, 2000; Jones & Callan, 2003; Miller & D'Esposito, 2005; Ojanen et al., 2005; Pekkola et al., 2006; Sekiyama, Kanno, Miura, & Sugita, 2003; Skipper, Goldin-Meadow, Nusbaum, & Small, 2007; Skipper, Nusbaum, & Small, 2005; Wright, Pelphrey, Allison, McKeown, & McCarthy, 2003). The extent to which the mechanism for audiovisual speech comprehension in the child compares to the adult case is unknown; in particular, it is unclear whether the neurobiological substrate in the developing brain incorporates additional regions or different patterns of effective connectivity. In this paper, we examine the potential developmental mechanisms that result in a mature system implementing audiovisual speech comprehension, and how this system changes developmentally in the interactions among brain regions involved in the production and perception of speech.

* Corresponding author. Address: Human Neuroscience Laboratory, Department of Neurology, The University of Chicago, 5841 S. Maryland Ave. MC-2030, Chicago, IL 60637, USA. Fax: +1 773 702 2482.
E-mail address: adick@uchicago.edu (A.S. Dick).

Empirical evidence suggests that speech perception in children is less influenced by visual speech information than in adults. For example, studies assessing the development of audiovisual speech perception using incongruent "McGurk" stimuli report an increase in the influence of visual speech information with age, with development perhaps continuing even as late as the 11th year (Hockley & Polka, 1994; Massaro, 1984; Massaro et al., 1986; McGurk & MacDonald, 1976; Sekiyama & Burnham, 2008; Wightman, Kistler, & Brungart, 2006). Several factors likely contribute to the neural development of audiovisual speech, including both general factors (e.g., development of selective attention, increasing myelination, and synaptic pruning) and more specific factors (e.g., learning of oral motor patterns for speech). Here we focus on the child's increasing personal experience perceiving and producing speech, in an effort to gain insight into how children integrate audiovisual information during everyday verbal communication.

There is evidence to suggest that audiovisual speech integration is a skill that is acquired by experience listening to and observing speech over an extended period of time. In adults, for example, the amount of experience with a second language affects audiovisual processing in that language: Native French speakers with only beginning and intermediate skills in English are less sensitive to visual cues indicating a particular English consonant than either monolingual English speakers or more advanced French/English bilinguals (Werker, Frost, & McGurk, 1992). The role of experience in audiovisual integration gains further support from the presence of cross-linguistic differences. In one such example, Japanese adults are less influenced by visual information than English adults (Sekiyama & Burnham, 2008; Sekiyama & Tohkura, 1993), a difference that begins to emerge between 6 and 8 years (Sekiyama & Burnham, 2008). These data suggest that developmental differences in audiovisual speech integration are moderated by everyday perceptual experience with language.

Experience with speech *production* also contributes to the development of audiovisual speech comprehension. For example, children with articulatory difficulties perceive incongruent audiovisual syllables more often by their auditory component than do children without articulatory difficulties, who more often hear the fused percept or perceive the visual component alone (e.g., when presented with an auditory /pa/ and a visual /ka/, children tend to report hearing /pa/ instead of a fused percept /ta/, or visual /ka/; Desjardins, Rogers, & Werker, 1997, but see Dodd, McIntosh, Erdener, & Burnham, 2008). In addition, both groups of children are less likely to integrate the auditory and visual information into a fused percept, or to perceive the visual component alone, than adults. Further, children with cochlear implants who produce more intelligible speech demonstrate an improved ability to use visual speech information (Bergeson, Pisoni, & Davis, 2005; Lachs, Pisoni, & Kirk, 2001). Taken together, these findings suggest a relationship between speech production skill and audiovisual speech perception.

The development of audiovisual speech comprehension thus appears to involve mechanisms that relate visual speech information to articulatory speech representations, both of which are acquired through experience with one's native language (cf. Desjardins et al., 1997; Kuhl & Meltzoff, 1982, 1984; Sekiyama & Burnham, 2008). Specifically, we propose that as children experience the auditory, somatosensory, and motor consequences of produced speech sounds in their own speech and in others' speech, they develop a mapping system between sensory and motor output. This mapping allows for these components of the audiovisual speech signal to have a "predictive value" for each of the other components. For example, several authors have suggested that motor-speech representations constrain the interpretation of the incoming auditory signal (Callan et al., 2004; Skipper et al., 2005; Skipper, Nusbaum, & Small, 2006; Skipper, van Wassenhove, Nusbaum, & Small, 2007; van Wassenhove, Grant, & Poeppel, 2005;

Wilson & Iacoboni, 2006). In one model, visible articulatory movements of the speaker's lips and mouth invoke articulatory representations of the listener that could have generated the observed speech movements (Skipper, van Wassenhove, et al., 2007; van Wassenhove et al., 2005). These representations, based in prior articulatory experience, provide a set of possible phonetic targets to constrain the final interpretation of the speech sound (i.e., the visual information provides a "forward model" of the speech sound). Such motor-speech models draw on the listener's articulatory repertoire, and we argue that, because adults have more experience producing and perceiving speech than children, they have more precise predictors of the target speech sound.

As mentioned at the outset, the neural substrate of audiovisual speech perception consists of a widespread network of interconnected cortical regions. In general, brain networks develop though increasing integration among the component regions that define the network (Church et al., 2009; Fair, Cohen, et al., 2007; Fair, Dosenbach, et al., 2007; Karunanayaka et al., 2007). The primary objective of the present study was to characterize this developmental change for audiovisual speech comprehension. To do so we used structural equation modeling (SEM) to assess differences between adults and children in effective connectivity among left hemisphere brain regions important for language production and perception. Physiological studies have suggested that interactions among left IFGOp/PMv, SMG, STGp, PTe, and STSp support audiovisual speech perception (see Campbell, 2008 for review). In particular, the development of audiovisual speech might depend on interactions between inferior frontal/ventral premotor regions, and posterior temporal/inferior parietal regions. This pathway has been postulated to help relate motor (articulatory) and sensory (auditory and somatosensory) information about the identity of the speech target (Callan et al., 2004; Skipper et al., 2005; Skipper et al., 2006; Skipper, van Wassenhove, et al., 2007; van Wassenhove et al., 2005; Wilson & Iacoboni, 2006). Because adults have more experience both perceiving and producing speech, their sensory and motor repertoires are richer, and will have greater predictive value. Thus, we predict significant age differences in effective connectivity for audiovisual speech between inferior frontal/ventral premotor regions, and posterior temporal and inferior parietal regions.

In the current study, we used functional magnetic resonance imaging (fMRI) and SEM to study 24 adults and nine children during auditory-alone and audiovisual story comprehension. We compared effective connectivity between children and adults across the network, with particular attention to connectivity between temporal/parietal and frontal regions.

2. Materials and methods

2.1. Participants

Twenty-four adults (12 females, *M* age = 23.0 years, *SD* = 5.6 years) and nine children (7 females, range = 8–11 years, *M* age = 9.5 years, *SD* = 0.9 years) participated. Eight years was the youngest age in the available cohort, and in previous studies, development of audiovisual speech perception has been shown to occur in this age range, with few age differences beyond age 11 (Hockley & Polka, 1994; Massaro et al., 1986; McGurk & MacDonald, 1976; Sekiyama & Burnham, 2008; Wightman et al., 2006). Only one additional child was excluded for excessive motion (>1 mm). The adult sample was part of a prior investigation conducted in our laboratory, which used the same experimental manipulation (Dick, Goldin-Meadow, Hasson, Skipper, & Small, 2009; Skipper, Goldin-Meadow, et al., 2007; Skipper, Goldin-Meadow, Nusbaum, & Small, 2009). All participants were right-handed according to the Edinburgh handedness inventory (Oldfield, 1971),

had normal hearing (self-reported) and normal (or corrected to normal) vision and were native speakers of English. No participant had a history of neurological or psychiatric illness. All adult participants gave written informed consent. Participants under 18 years gave assent and informed consent was obtained from a parent. The Institutional Review Board of the Biological Sciences Division of The University of Chicago approved the study.

2.2. Stimuli

We used a passive listening paradigm to avoid explicit motor responses, which could introduce a confound in motor areas responsible for preparing and producing an action to respond (Small & Nusbaum, 2004; Yarkoni, Speer, Balota, McAvoy, & Zacks, 2008). Participants were instructed to watch and listen to short adaptations of Aesop's Fables ($M = 53s$; $SD = 3s$) that were presented with and without visual speech information. Although the overall study incorporated four story-telling conditions, only two are included in the present analysis, *Audiovisual* (AV; with face information but without manual gestures) and *Auditory* (A; with no visual input). Each participant heard one AV and one A story in each of two runs for a total of two stories of each type. The stories were separated by a 16 s *Baseline* fixation condition. In the AV condition, participants watched and listened to a female storyteller who rested her hands in her lap. She was framed from waist to head with sufficient width to allow full perception of the upper body. In the A condition, participants listened to the stories while watching a fixation cross presented on the screen. Audio was delivered at a sound pressure level of 85 dB-SPL through MRI-compatible headphones (Resonance Technologies, Inc., Northridge, CA). Video stimuli were viewed through a mirror attached to the head coil. Following each run, participants responded to true/false questions about each story using a button box. Mean accuracy was high for both adults (AV $M = 87\%$; A $M = 82\%$) and children (AV $M = 84\%$; A $M = 84\%$), with no significant group or condition differences or interaction (all t's < 1). These results suggest both children and adults paid attention to the stories in both conditions.

2.3. Data collection

MRI scans were acquired at 3-Tesla with a standard quadrature head coil (General Electric, Milwaukee, WI, USA). Volumetric T1-weighted scans (120 axial slices, $1.5 \times .938 \times .938$ mm resolution) provided high-resolution anatomical images. For the functional scans, thirty slices (voxel size $5.00 \times 3.75 \times 3.75$ mm) were acquired in the sagittal plane using spiral blood oxygen level dependent (BOLD) acquisition (TR/TE = 2000 ms/25 ms, FA = 77°; Noll, Cohen, Meyer, & Schneider, 1995). The first four BOLD scans of each of the two runs were discarded to avoid images acquired before the signal reached a steady state.

2.4. Scanning children

Special steps were taken to ensure that children were properly acclimated to the scanner environment. Following Byars et al. (2002), we included a "mock" scan during which children practiced lying still in the scanner while listening to prerecorded scanner noise. When children felt confident to enter the real scanner, the session began.

2.5. Data analysis I: preprocessing

Two analyses were performed: a "block" analysis correlating the hemodynamic response during story presentation to an extrinsic hemodynamic response function, and a network analysis using structural equation modeling (SEM). The following steps were implemented:

2.5.1. Preprocessing

Preprocessing steps were conducted using Analysis of Functional Neuroimages software (AFNI; http://afni.nimh.nih.gov) on the native MRI images. For each participant, image processing consisted of (1) three-dimensional motion correction using weighted least-squares alignment of three translational and three rotational parameters, and registration to the first non-discarded image of the first functional run, and to the anatomical volumes; (2) despiking and mean normalization of the time series; (3) inspection and censoring of time points occurring during excessive motion (>1 mm; Johnstone et al., 2006); (4) modeling of sustained hemodynamic activity within a story via regressors corresponding to the conditions, convolved with a gamma function model of the hemodynamic response derived from Cohen (1997). We also included linear and quadratic drift trends, and six motion parameters obtained from the spatial alignment procedure. This analysis resulted in regression coefficients (beta weights) and associated t statistics measuring the reliability of the coefficients. False Discovery Rate (FDR; Benjamini & Hochberg, 1995; Genovese, Lazar, & Nichols, 2002) statistics were calculated for each beta value; (5) to remove additional sources of spurious variance unlikely to represent signal of interest, we regressed from the time series signal from both lateral ventricles, and from bilateral white matter (Fox et al., 2005).

2.5.2. Time series assessment and temporal re-sampling in preparation for SEM

Due to counterbalancing, story conditions differed slightly in length (i.e., across participants the same stories were used in different conditions). Because SEM analyzes covariance structures, the time series must be the same length across individuals. To standardize time series length, we first imported time series from significant voxels ($p < .01$; FDR corrected) in predefined ROIs (see below), removed outlying voxels (>10% signal change), and averaged the signal to achieve a representative time series across the two runs for each ROI for each condition (AV and A; baseline time points were excluded). We then re-sampled these averaged time series down to 92 s using a locally weighted scatterplot smoothing (LOESS) method. In this method each re-sampled data point is estimated with a weighted least squares function, giving greater weight to actual time points near the point being estimated, and less weight to points farther away (Cleveland & Devlin, 1988). Non-significant Box's M tests indicated no differences in the variance–covariance structure of the re-sampled and original data. The SEM analysis was conducted on these re-sampled time series.

2.6. Data analysis II: standard analysis of activation differences

We conducted second-level group analysis on a two-dimensional surface rendering of the brain constructed in Freesurfer (http://surfer.nmr.mgh.harvard.edu; Dale, Fischl, & Sereno, 1999; Fischl, Sereno, & Dale, 1999). Note that although children and adults do show differences in brain morphology, Freesurfer has been used to successfully create surface representations for children (Tamnes et al., 2009), and even neonates (Pienaar, Fischl, Caviness, Makris, & Grant, 2008). Further, in the age range we investigate here, atlas transformations similar to the kind used by Freesurfer have been shown to lead to robust results without errors when comparing children and adult functional images (Burgund et al., 2002; Kang, Burgund, Lugar, Petersen, & Schlaggar, 2003). Using AFNI, we interpolated regression coefficients, representing percent signal change, to specific vertices on the surface representation of the individual's brain. Image registration across the group required an additional standardization step accomplished with icosahedral tessellation and projection (Argall, Saad, & Beauchamp, 2006). The functional data were smoothed on the surface (4 mm FWHM) and imported to a MySQL relational

database (http://www.mysql.com/). The R statistical package (version 2.6.2; http://www.R-project.org) was then used to query the database and analyze the information stored in these tables (for details see Hasson, Skipper, Wilde, Nusbaum, & Small, 2008). Finally, we created an average of the cortical surfaces in Freesurfer on which to display the results of the whole-brain analysis.

We conducted a mixed (fixed and random) effects Condition (repeated measure; 2) × Age Group (2) × Participant (33) ANOVA on a vertex-by-vertex basis using the normalized regression coefficients as the dependent variable. We assessed comparisons with the resting baseline, between age groups and conditions, and 2 (Age) × 2 (Condition) interaction contrasts. We also removed statistical outliers (>3 SDs from the mean of transverse temporal gyrus; outliers represented <1% of the data). To control for the family-wise error (FWE) rate given multiple comparisons, we clustered the data using a non-parametric permutation method. This method proceeds by re-sampling the data under the null hypothesis without replacement, making no assumptions about the distribution of the parameter in question (see Hayasaka & Nichols, 2003; Nichols & Holmes, 2002 for implementation details). Using this method, we determined a minimum cluster size (e.g., taking cluster sizes above the 95th percentile of the random distribution controls for the FWE at the $p < .05$ level). Reported clusters used a per-surface-vertex threshold of $p < .01$ and controlled for the FWE rate of $p < .05$.

2.6.1. Signal-to-noise ratio and analysis

We carried out a signal-to-noise ratio (SNR) analysis to determine if there were any cortical regions where, across participants and groups, it would be impossible to find experimental effects simply due to high noise levels (see Parrish, Gitelman, LaBar, & Mesulam, 2000 for rationale of using this method in fMRI studies). We present the details of these analyses in the Supplementary materials and report the results below.

2.7. Data analysis III: network analysis using structural equation modeling

The SEM analysis was performed using AMOS (Arbuckle, 1989). Where applicable, we followed the procedural steps in Solodkin, Hlustik, Chen, and Small (2004) and McIntosh and Gonzalez-Lima (1994), and report those steps here in abbreviated form.

2.7.1. Specification of theoretical anatomical model

The specification of a theoretical anatomical model requires the definition of the nodes of the network and the directional connections (i.e., paths) among them. Our theoretical model represents a compromise between the complexity of the neural systems implementing language comprehension and the interpretability of the resulting model. Although a complex model might account for most or all known anatomical connections, it would be nearly impossible to interpret (McIntosh & Gonzalez-Lima, 1992; McIntosh et al., 1994). Further, our hypotheses are focused on left hemisphere frontal, temporal, and parietal regions. Thus, we specified a left hemisphere theoretical network for language comprehension, which included ten brain regions (ROIs) and their connections. These regions were chosen based on the results of recent functional imaging findings examining discourse comprehension in adults (Hasson, Nusbaum, & Small, 2007; Skipper et al., 2005) and in children (Karunanayaka et al., 2007; Schmithorst, Holland, & Plante, 2006). Connectivity among the regions was constrained by known anatomical connectivity in macaques (Schmahmann & Pandya, 2006). ROIs were defined on each individual surface representation using an automated parcellation procedure in Freesurfer (Desikan et al., 2006; Fischl et al., 2004), incorporating the neuroanatomical conventions of Duvernoy (1991). We manually edited the default parcellation to delineate anterior and posterior portions

of the predefined temporal regions, and dorsal and ventral portions of the predefined premotor region. The specific anatomical regions are described in Table 1. Surface interpolation of functional data inherently results in spatial smoothing across contiguous ROIs (and potentially spurious covariance). To avoid this, surface ROIs were imported to the native MRI space for SEM, and the time series were not spatially smoothed.

We next proceeded to construction of the structural equation models, which required the following steps:

2.7.2. Generation of the covariance matrix

For each age group (children and adults), we generated a variance–covariance matrix based on the mean time series from active voxels ($p < .01$, FDR corrected) across all participants, for all ROIs. One covariance matrix per group, per condition was generated.

2.7.3. Generation of and solving of structural equations

Initial models were constructed in AMOS, which uses an iterative maximum likelihood method to obtain a solution for each path coefficient representing the effective connectivity between each node (i.e., anatomical ROI) of the network. The best solution of the set of equations minimizes the difference between the observed and predicted covariance matrices.

2.7.4. Goodness of fit between the predicted and observed variance–covariance matrices

Model fit was assessed against a χ^2 distribution with $q(q + 1)/2 - p$ degrees of freedom (where q is the number of nodes of the network, and p is the number of unknown coefficients). Good model fit is obtained if the null hypothesis (specifying no difference between the observed and predicted covariance matrices) is not rejected (Barrett, 2007).

2.7.5. Comparison between models

Multiple group comparison was used to assess model differences between (1) AV and A within age group, and (2) adult and child within condition. Differences between the models were assessed in AMOS using the "stacked model" approach (McIntosh & Gonzalez-Lima, 1994), according to which both groups are simultaneously fit to the same model, with the null hypothesis that the path coefficients between the groups do not differ (null model). In the alternative model, paths of interest are allowed to differ. Significant differences in the χ^2 fit for both models are assessed with reference to a critical χ^2 value $\chi^2_{crit,df=1} = 3.84$. A significant difference implies that better model fit is achieved when the paths are allowed to vary across groups. Although the approach requires identical models, there were fewer active areas for A than for AV. Data from these missing visual and fusiform nodes were obtained by generating a random time series vector, with the missing path coefficients to and from these regions fixed to a value of zero.

3. Results

3.1. Signal-to-noise ratio

Simulations indicated that in the current design, the minimum SNR needed to detect a signal change of 0.5% was 54, and that needed to detect a signal change of 1% was 27 (see Supplementary materials). We analyzed the mean SNR across participants from 58 cortical and subcortical anatomical ROIs. In the regions examined, mean SNR ranged from a low of 13.8 (SD = 7.1) in the right temporal pole (a region of high susceptibility artifact) to a high of 134.3 (SD = 25.6) in the right superior precentral sulcus. Importantly, in most regions SNR was sufficient to detect even small signal changes, and on average SNR did not significantly differ across

Table 1
Anatomical description of the cortical regions of interest.

ROI	Anatomical structure	Brodmann's area	Delimiting landmarks
IFGTr	*pars triangularis* of the inferior frontal gyrus	45	A = A coronal plane defined as the rostral end of the anterior horizontal ramus of the sylvian fissure P = Vertical ramus of the sylvian fissure S = Inferior frontal sulcus I = Anterior horizontal ramus of the sylvian fissure
IFGOp/ PMv	*pars opercularis* of the inferior frontal gyrus, inferior precentral sulcus, and inferior precentral gyrus	6, 44	A = Anterior vertical ramus of the sylvian fissure P = Central sulcus S = Inferior frontal sulcus, extending a horizontal plane posteriorly across the precentral gyrus I = Anterior horizontal ramus of the sylvian fissure to the border with insular cortex
SMG	Supramarginal gyrus	40	A = Postcentral sulcus P = Sulcus intermedius primus of Jensen S = Intraparietal sulcus I = Sylvian fissure
STa	Anterior portion of the superior temporal gyrus and planum polare	22	A = Inferior circular sulcus of insula P = A vertical plane drawn from the anterior extent of the transverse temporal gyrus S = Anterior horizontal ramus of the sylvian fissure I = Dorsal aspect of the upper bank of the superior temporal sulcus
Auditory	Transverse temporal gyrus and sulcus (Heschl's gyrus and sulcus)	41, 42	A = Inferior circular sulcus of insula and planum polare P = Planum temporale M = Sylvian fissure L = Superior temporal gyrus
STp	Posterior portion of the superior temporal gyrus and planum temporale	22, 42	A = A vertical plane drawn from the anterior extent of the transverse temporal gyrus P = Angular gyrus S = Supramarginal gyrus I = Dorsal aspect of the upper bank of the superior temporal sulcus
MTGp	Posterior middle temporal gyrus	21	A = A vertical plane drawn from the anterior extent of the transverse temporal gyrus P = Temporo-occipital incisure S = Superior temporal sulcus I = Inferior temporal sulcus
STSp	Posterior superior temporal sulcus	22	A = A vertical plane drawn from the anterior extent of the transverse temporal gyrus P = Angular gyrus and middle occipital gyrus and sulcus S = Angular and superior temporal gyrus I = Middle temporal gyrus
Fusiform gyrus	Fusiform or lateral occipito-temporal gyrus	37	A = Anterior transverse collateral sulcus P = Posterior transverse collateral sulcus M = Medial occipito-temporal sulcus L = Lateral occipito-temporal sulcus
Visual	Inferior occipital gyrus and sulcus, middle occipital gyrus and sulcus, calcarine sulcus, and occipital pole	17, 18, 19	A = On the lateral surface, a line starting from the parieto-occipital fissure to the temporo-occipital incisure; the medial surface included calcarine sulcus extending to the precuneus P = Posterior occipital pole S = Superior occipital gyrus I = Lateral occipito-temporal gyrus

Note: A = Anterior; P = Posterior; S = Superior; I = Inferior; M = Medial; L = Lateral.

adults and children (Adult M = 66.3; SD = 29.0 vs. Child M = 62.0; SD = 36.4, $t(56)$ = 0.82, p > .05). This suggests sufficient SNR to detect differences within and across groups.

3.2. Auditory and audiovisual story comprehension compared to baseline

We first examined signal changes, both positive ("activations") and negative ("deactivations"), for each condition (AV, A) relative to a resting baseline, across groups. Both contrasts showed bilateral activation in frontal, inferior parietal, and temporal regions. An exception to this was a lack of right frontal and parietal activation in the child group for both conditions (see Fig. 1 and Table 2 for cluster size and stereotaxic coordinates). Additional activation in bilateral occipital–temporal regions was found in the AV condition. Deactivations were found in posterior cingulate, precuneus, cuneus, lateral superior parietal cortex, lingual gyrus, and, for chil-

dren, in the right superior frontal gyrus. These findings are comparable to prior studies of language comprehension in both adults and children (Hasson, Nusbaum, et al., 2007; Schmithorst et al., 2006; Wilson, Molnar-Szakacs, & Iacoboni, 2008).

3.3. Comparison of activation differences: audiovisual vs. auditory-only

For both adults and children, AV elicited greater BOLD signal intensity than A in bilateral occipital regions and posterior fusiform gyrus (Table 3). Additional differences were found (for adults) in left parahippocampal gyrus, and (for children) in bilateral thalamus, left inferior frontal gyrus, anterior insula, and bilateral posterior superior temporal gyrus (extending to posterior superior temporal sulcus in the left hemisphere). For adults, A was more active than AV in bilateral posterior cingulate gyrus, left lingual gyrus, and left parahippocampal gyrus. Similarly, for children, A
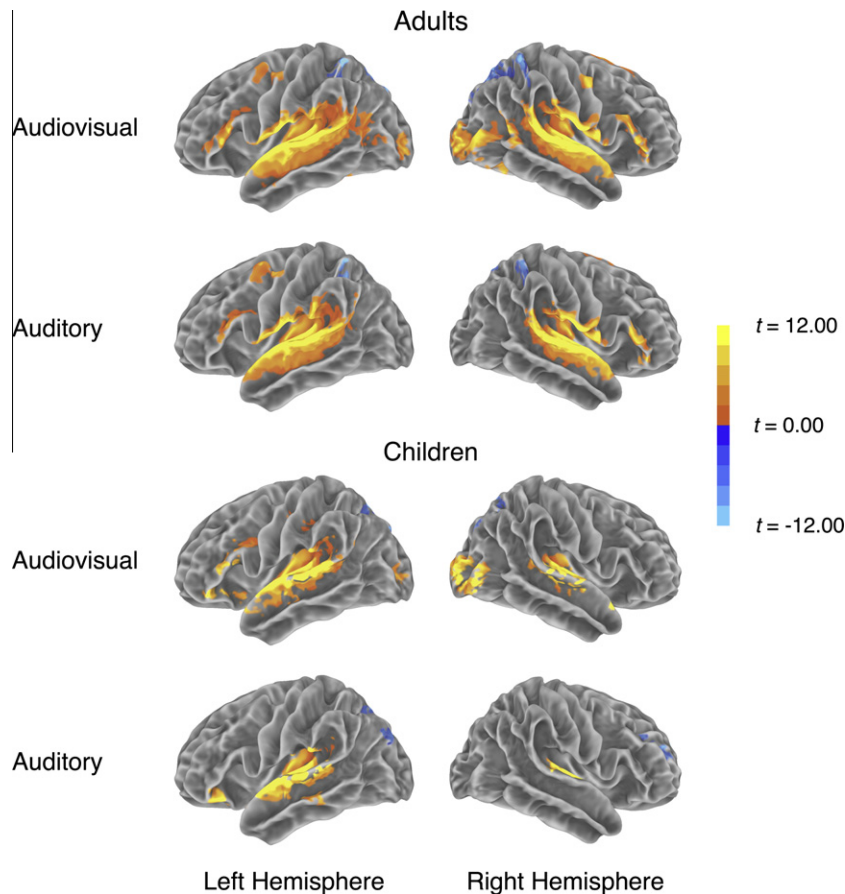
Fig. 1. Whole-brain analysis results for each condition compared to *Baseline* for both adults and children. The individual per-vertex threshold was $p < .01$ (corrected FWE $p < .05$).

was associated with stronger activation than AV in left posterior cingulate and left lingual gyrus, and right cuneus.

3.4. Comparison of activation differences: adults vs. children

Although no significant clusters showed greater BOLD intensity for adults than children in the AV condition, there were differences favoring adults in the A condition (Table 3). Right superior temporal gyrus, right middle frontal gyrus, left insula and left postcentral gyrus demonstrated greater activity for adults in the A condition. Children elicited greater activity in left medial frontal gyrus and precuneus/posterior cingulate during the AV condition, but no significant differences favored children in the A condition. Finally, interaction contrasts (i.e., Adults [AV–A] – Children [AV–A] ≠ 0; Jaccard, 1998) conducted over the whole brain revealed no clusters that met the FWE correction for multiple comparisons.

3.5. Structural equation modeling analysis of base networks and network differences

3.5.1. Goodness of fit of base network models

Good fit was obtained across both conditions and both age groups (Adult AV: $\chi^2 = 26.77$, $df = 17$, $p > .06$; Adult A: $\chi^2 = 34.07$, $df = 23$, $p > .06$; Children AV: $\chi^2 = 23.50$, $df = 17$, $p > .13$; Children A: $\chi^2 = 34.36$, $df = 23$, $p > .06$), indicating that the hypothesized model should be retained. Squared multiple correlations revealed that many regions consistently showed a moderate to high degree of explained variance across groups and conditions (i.e., between 40 and 80% variance explained; these included STa, STp, STSp,

MTGp). In three regions (SMG, IFGTr, IFGOp/PMv), however, explained variance was more variable across groups and conditions, explaining between 1% and 72% of the variance. This indicates not only potential influence on these regions from brain areas absent in the network model, but also greater influence of the task demands on effective connectivity to and from these regions, leading to greater variability across conditions and groups.

The theoretical networks incorporating all regions are provided in Fig. 2. As expected, both networks revealed moderate to strong connectivity from auditory cortex to anterior and posterior temporal cortices. The AV network additionally revealed moderate to strong connectivity from visual cortex and fusiform gyrus to these temporal regions.

3.5.2. Analysis of network differences

Comparison of networks proceeded in two steps using the multiple group or "stacked model" approach (McIntosh & Gonzalez-Lima, 1994). We first compared the broader 10-node network across groups and across conditions. This analysis yielded significant differences in the pattern of network interactions between age groups and across conditions. This suggests that, although the base models shared the same nodes and predicted connectivity, and provided a good fit for both groups and conditions, there were significant differences in the connection weights. For both age groups, there were differences between the networks for AV and A (AV vs. A for Adults: $\chi^2 = 186.3$, $df = 68$, $p < .001$; AV vs. A for Children: $\chi^2 = 104.7$, $df = 68$, $p < .01$). In addition, there were differences across age for both conditions (Adults vs. Children for the AV condition: $\chi^2 = 132.0$, $df = 62$, $p < .001$; Adults vs. Children for the A condition: $\chi^2 = 131.0$, $df = 68$, $p < .05$).

Table 2
Regions showing reliable activation or deactivation relative to baseline, for both conditions across both groups.

Region	Adults								Children							
	Audiovisual				Auditory-only				Audiovisual				Auditory-only			
	Talairach			CS	Talairach			CS	Talairach			CS	Talairach			CS
	x	y	z		x	y	z		x	y	z		x	y	z	
Activation																
Occipital																
L. Mid occipital gyrus	−29	−91	3	313	–	–	–	–	−32	−88	3	379	–	–	–	–
R. Mid occipital gyrus	40	−81	3	1325	–	–	–	–	35	−87	0	1889	–	–	–	–
Temporal and parietal																
L. Angular gyrus	–	–	–	–	–	–	–	–	−35	−54	38	419	–	–	–	–
L. Fusiform	−41	−46	−16	522	–	–	–	–	–	–	–	–	–	–	–	–
L. Mid temporal gyrus	–	–	–	–	–	–	–	–	–	–	–	–	−61	−30	−9	387
L. Parahipp gyrus	−25	−6	−12	284	–	–	–	–	–	–	–	–	–	–	–	–
L. Parahipp gyrus	–	–	–	–	–	–	–	–	−27	−12	−12	380	–	–	–	–
L. Postcentral gyrus	–	–	–	–	–	–	–	–	−49	−16	33	663	–	–	–	–
L. Postcentral gyrus	–	–	–	–	–	–	–	–	−48	−28	37	320	–	–	–	–
L. Sup temporal gyrus	−51	−34	9	17,006	−52	−28	6	14,644	−49	−22	8	10,111	−50	−19	7	6880
R. Fusiform	42	−58	−14	350	–	–	–	–	–	–	–	–	–	–	–	–
R. Mid temporal gyrus	54	−56	13	576	–	–	–	–	–	–	–	–	–	–	–	–
R. Sup temporal gyrus	–	–	–	–	55	−21	6	11,068	51	−22	6	4736	52	−19	6	1469
R. Trans temp gyrus	47	−22	10	12,288	–	–	–	–	–	–	–	–	–	–	–	–
Frontal																
L. Inf frontal gyrus	−50	25	15	757	−46	22	21	602	−42	15	24	618	−44	33	−9	340
L. Inf frontal gyrus	–	–	–	–	–	–	–	–	−39	29	−8	532	–	–	–	–
L. Inf frontal gyrus	–	–	–	–	–	–	–	–	−52	27	12	448	–	–	–	–
L. Middle frontal gyrus	–	–	–	–	−41	−3	50	782	–	–	–	–	–	–	–	–
L. Precentral gyrus	−48	−6	49	303	–	–	–	–	–	–	–	–	–	–	–	–
R. Inf frontal gyrus	50	29	4	1349	55	30	7	552	–	–	–	–	–	–	–	–
R. Precentral gyrus	54	−2	45	385	–	–	–	–	–	–	–	–	–	–	–	–
R. Precentral gyrus	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
R. Sup frontal gyrus	8	12	65	385	9	10	68	302	–	–	–	–	–	–	–	–
Deactivation																
Occipital and parietal																
L. Cuneus	−12	−75	12	823	–	–	–	–	–	–	–	–	–	–	–	–
L. Cuneus	−24	−79	25	426	–	–	–	–	–	–	–	–	–	–	–	–
L. Precuneus	−6	−68	41	1490	−6	−64	37	281	−10	−66	29	4469	−8	−66	45	1901
L. Precuneus	–	–	–	–	–	–	–	–	−23	−77	29	730	−14	−84	38	307
L. Sup occipital gyrus	–	–	–	–	–	–	–	–	–	–	–	–	−36	−82	33	534
R. Cuneus	13	−74	12	998	–	–	–	–	9	−74	16	703	–	–	–	–
R. Precuneus	11	−50	41	4880	–	–	–	–	20	−74	45	2284	10	−71	47	785
R. Precuneus	–	–	–	–	–	–	–	–	8	−62	19	300	12	−58	30	378
Temporal and parietal																
L. Fusiform	−23	−61	−8	666	–	–	–	–	–	–	–	–	−32	−62	−15	282
L. Insula	–	–	–	–	–	–	–	–	–	–	–	–	−28	22	9	313
L. Lingual gyrus	–	–	–	–	–	–	–	–	−19	−62	−6	744	–	–	–	–
L. Posterior cingulate	−4	−29	31	1332	−3	−22	34	1246	−5	−27	27	1462	−6	−24	31	1454
L. Superior parietal	−32	−49	50	425	–	–	–	–	–	–	–	–	–	–	–	–
R. Lingual gyrus	25	−65	−4	643	–	–	–	–	–	–	–	–	–	–	–	–
R. Posterior cingulate	–	–	–	–	5	−29	33	1946	7	−33	29	1957	7	−30	30	1725
R. Superior parietal	36	−46	55	285	10	−71	56	501	–	–	–	–	–	–	–	–
Frontal																
R. Sup frontal gyrus	–	–	–	–	–	–	–	–	–	–	–	–	25	46	21	530

Note: Individual voxel threshold *p* < .01, corrected (FWE *p* < .05). Center of mass defined by Talairach and Tournoux coordinates in the volume space. CS = Cluster size in number of surface vertices. L = Left. R = Right. Inf = Inferior. Mid = Middle. Parahipp = Parahippocampal. Sup = Superior.

Table 3
Regions showing reliable differences across conditions and across groups.

Region	Talairach			BA	CS	MI
	x	y	z			
Adults: AV > A						
L. Middle occipital gyrus	−28	−93	1	18	453	0.63
L. Parahippocampal gyrus	−14	3	−12	34	329	0.63
R. Fusiform	41	−54	−12	37	469	0.67
R. Middle occipital gyrus	41	−82	2	18	1643	0.84
Adults: A > AV						
L. Lingual gyrus	−17	−66	1	19	349	0.40
L. Parahippocampal gyrus	−20	−51	−6	19	908	0.32
L. Posterior cingulate	−11	−59	14	30	454	0.34
R. Posterior cingulate	14	−55	12	30	597	0.40
Children: AV > A						
L. Fusiform gyrus	−37	−69	−16	19	410	1.39
L. Inferior frontal gyrus	−29	15	−11	47	375	0.65
L. Insula	−39	10	12	13	319	0.34
L. Insula	−36	−5	−9	–	286	0.34
L. Middle occipital gyrus	−39	−82	10	19	594	0.88
L. Superior temporal gyrus	−54	−54	17	22/39	320	0.44
L. Thalamus	−7	−22	10	–	1092	0.80
R. Middle occipital gyrus	42	−81	4	18/19	1232	0.88
R. Superior temporal gyrus	54	−37	10	22	490	0.71
R. Superior temporal gyrus	52	−31	12	42	480	0.61
R. Superior temporal gyrus	46	−57	17	22	367	0.57
R. Thalamus	9	−20	6	–	553	0.59
Children: A > AV						
L. Lingual gyrus	−11	−63	−3	18/19	454	0.69
L. Posterior cingulate	−18	−61	10	30	502	0.46
R. Cuneus	5	−77	12	17	418	0.67
Audiovisual: adults > children						
No significant clusters	–	–				
Audiovisual: children > adults						
L. Medial frontal gyrus	−7	44	30	9	272	0.15
L. Precuneus	−9	−59	41	7	521	0.25
Auditory: adults > children						
L. Insula	−33	−33	22	13	315	1.01
L. Postcentral gyrus	−56	−10	14	43	340	1.60
R. Middle frontal gyrus	24	51	20	10	302	0.19
R. Superior temporal gyrus	62	−36	10	22	894	1.43
Auditory: children > adults						
No significant clusters	–	–	–			

Note: Individual voxel threshold $p < .01$, corrected (FWE $p < .05$). Center of mass defined by Talairach and Tournoux coordinates in the volume space. BA = Brodmann area. CS = Cluster size in number of surface vertices. L = Left. R = Right. MI = Maximum intensity (in terms of percent signal change).

Based on our a priori hypotheses, we focused our analysis of differences for specific pathways on the relationships among IFGOp/PMv, SMG, STp, and STSp (see Petersson, Reis, Askelöf, Castro-Caldas, & Ingvar, 2000 for a similar approach analyzing sub-networks within broader networks; shown as red nodes in Fig. 3). In this way we attempted to account for a sufficient degree of observed covariance by establishing the structure of the broader 10-node network, while focusing on our questions of interest within a smaller sub-network comprised of four nodes. Seven path coefficients within each network, comprising the interactions among only the four nodes of interest, were tested for significant differences (Fig. 3 and Table 4).

We first compared differences across age for both conditions. Comparing Adults to Children within the AV condition, we found that only one connection within this sub-network, IFGOp/PMv → SMG, was significantly different ($\chi^2_{\text{diff},1} = 5.7$, $p < .05$). When a similar comparison was conducted for the A condition, two age-related differences were found. The influence of IFGOp/PMv on STp ($\chi^2_{\text{diff},1} = 11.1$, $p < .001$), and the influence of STSp on STp ($\chi^2_{\text{diff},1} = 8.5$, $p < .05$) differed across groups.

We next examined condition differences within each age group. The effective connectivity of IFGOp/PMv on SMG, which showed an

age difference for AV, also differed when we compared AV to A within children ($\chi^2_{\text{diff},1} = 4.8$, $p < .05$). An additional significant difference was for the influence of IFGOp/PMv on STSp ($\chi^2_{\text{diff},1} = 7.90$, $p < .01$). Finally, several differences between the AV and A condition were found for adults. Effective connectivity of STSp on STp and SMG was stronger for AV than for A ($\chi^2_{\text{diff},1} = 6.0$, $p < .05$; ($\chi^2_{\text{diff},1} = 10.4$, $p < .001$). The influence of SMG on STp was stronger for A than for AV ($\chi^2_{\text{diff},1} = 7.0$, $p < .01$). The influence of IFGOp/PMv on STSp also differed across conditions ($\chi^2_{\text{diff},1} = 5.7$, $p < .05$). The magnitudes of the path coefficients are presented in Table 4. We elaborate on these findings below.

4. Discussion

We investigated age-related differences in the neurobiological substrates of audiovisual speech using a network modeling approach. We had expected differences in sensitivity to visual speech information manifested by differences in network connectivity among brain regions important for audiovisual speech comprehension. Consistent with prior network-level investigations of the development of story comprehension (Karunanayaka et al., 2007), clear differences were found in the functional relationships among regions of a fronto-temporo-parietal network across age groups and conditions. The principal finding was that, for audiovisual but not auditory-only comprehension, the influence of IFGOp/PMv on SMG differed across age groups, and, in children, between audiovisual and auditory-only comprehension. These results suggest that the development of audiovisual speech comprehension proceeds through changes in the functional interactions among brain regions involved in both language production and perception.

4.1. Development of a left fronto-temporo-parietal network for audiovisual language comprehension

Relative to a resting baseline, both children and adults activated a similar network of brain regions in response to audiovisual speech. When we investigated age differences in BOLD signal intensity in the AV condition, we found no significant clusters showing greater BOLD intensity for adults. We also found no evidence that condition modulated age differences in BOLD signal intensity (i.e., there was no condition by age interaction). Compared to adults, however, significant clusters showing greater BOLD intensity for children during AV were found in left precuneus and medial frontal gyrus. These medial prefrontal and posterior midline regions are commonly associated with a putative "default network" consisting of brain regions that reliably show decreased activation during the performance of an exogenous cognitive task (Fox et al., 2005; Shulman et al., 1997). Further, the degree of activation in these regions has been associated with encoding linguistic stimuli to memory, in some cases showing that greater deactivation is correlated with better retention, possibly reflecting the more efficient allocation of resources to the cognitive task (Clark & Wagner, 2003; Daselaar, Prince, & Cabeza, 2004). It is possible that the difference in activity in these regions reflects less efficient processing for children during AV speech, potentially because the added visual information is more helpful for adults than for children.

It is important to note, though, that no significant differences in BOLD signal intensity favoring either age group were found in brain regions previously associated with processing audiovisual speech. Age-related differences for these regions were only found in the SEM analysis of effective connectivity. Specifically, although no age differences were found in the comprehension of the stories, the network analysis showed that the networks for audiovisual
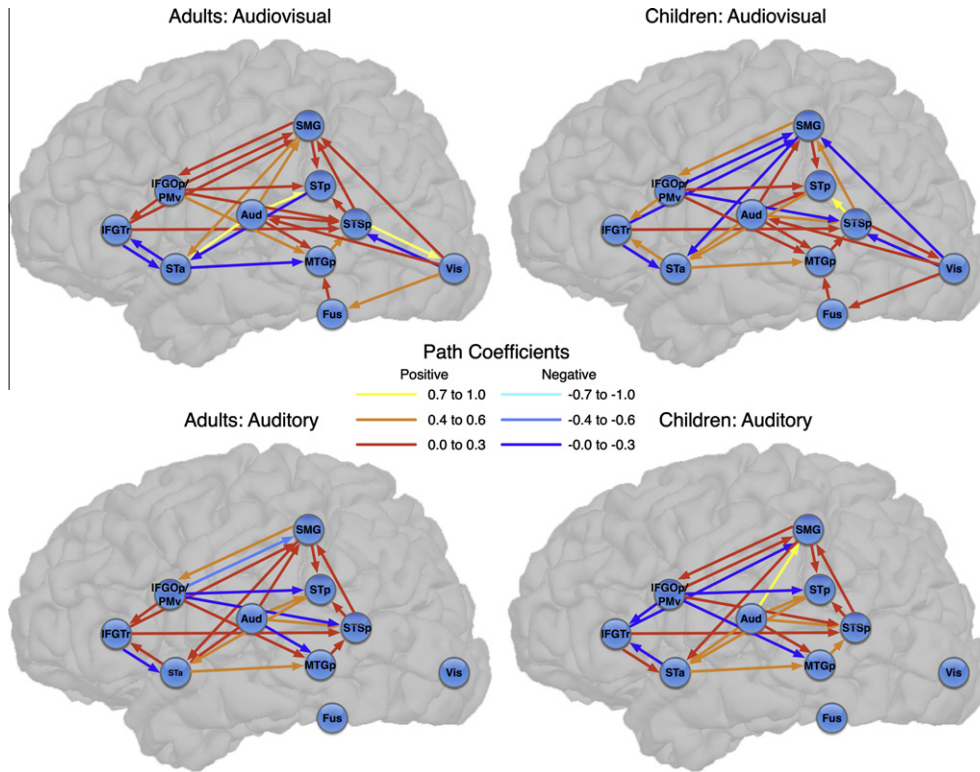
Fig. 2. Base theoretical network models for each condition and for each age group. Path coefficients represent standardized values. For the auditory-only networks, path coefficients connecting visual and fusiform nodes to the rest of the network were set to zero. IFGTr = Inferior frontal gyrus (*pars triangularis*); IFGOp/PMv = Inferior frontal gyrus (*pars opercularis*)/ventral premotor; SMG = Supramarginal gyrus; STp = Posterior superior temporal (posterior superior temporal gyrus and planum temporale); Aud = Auditory (transverse temporal gyrus and sulcus); STa = Anterior superior temporal (anterior superior temporal gyrus and planum polare); STSp = Posterior superior temporal sulcus; MTGp = Posterior middle temporal gyrus; Fus = Fusiform gyrus; Vis = Visual (striate and extrastriate cortex).
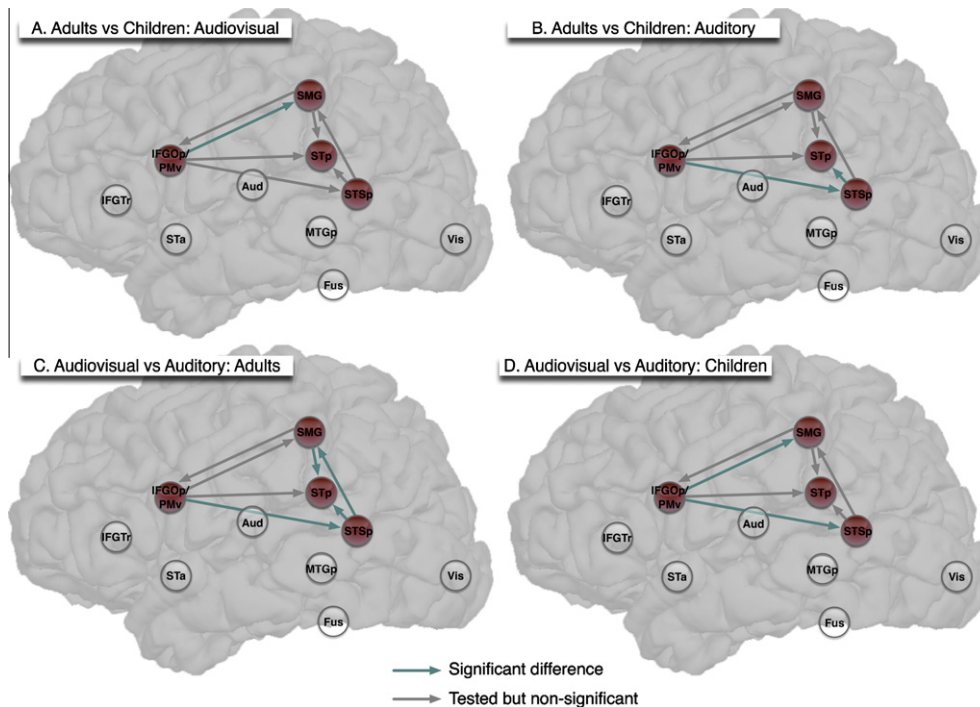
Fig. 3. Analysis of age and condition differences within the sub-network of interest. Red nodes were identified as important for connecting auditory and visual information during speech comprehension. Clear nodes comprised the remaining nodes of the theoretical network. Green arrows indicate a significant difference for that pathway for the comparison of interest. Grey arrows indicate pathways that were assessed for significance for the comparison of interest, but indicated no significant differences. IFGTr = Inferior frontal gyrus (*pars triangularis*); IFGOp/PMv = Inferior frontal gyrus (*pars opercularis*)/ventral premotor; SMG = Supramarginal gyrus; STp = Posterior superior temporal (posterior superior temporal gyrus and planum temporale); Aud = Auditory (transverse temporal gyrus and sulcus); STa = Anterior superior temporal (anterior superior temporal gyrus and planum polare); STSp = Posterior superior temporal sulcus; MTGp = Posterior middle temporal gyrus; Fus = Fusiform gyrus; Vis = Visual (striate and extrastriate cortex).

Table 4
Standardized path coefficients for significant comparisons of interest.

	Adults vs. Children		
	Network path	Adults	Children
Audiovisual			
	IFGOp/PMv → SMG	.10	−.24
Auditory			
	STSp → STp	.13	.36
	IFGOp/PMv → STSp	−.19	.23
	Audiovisual vs. Auditory		
	Network path	Audiovisual	Auditory
Adults			
	SMG → STp	.13	.19
	STSp → STp	.25	.13
	STSp → SMG	.28	.09
	IFGOp/PMv → STSp	.09	−.19
Children			
	IFGOp/PMv → SMG	−.24	.23
	IFGOp/PMv → STSp	−.11	.23

Note: IFGOp/PMv = Inferior frontal gyrus (*pars opercularis*)/ventral premotor; SMG = Supramarginal gyrus; STp = Posterior superior temporal (posterior superior temporal gyrus and planum temporale); STSp = Posterior superior temporal sulcus.

speech differed across age group, and across conditions (AV and A) within each age group. Further, both condition and age group modulated the influence of IFGOp/PMv on inferior parietal and posterior temporal brain regions. When visual speech information was present, the influence of IFGOp/PMv on SMG and STSp was moderately positive for adults (Fig. 3A and C), but moderately negative for children (Figs. 3A and D), though a significant age difference was found only for the influence of IFGOp/PMv on SMG (Fig. 3A). The connection weights here indicated that greater activity in IFGOp/PMv predicted greater activity in SMG for adults, but predicted reduced activity in SMG for children. For children, this pathway also differed between AV and A conditions, with greater activity in IFGOp/PMv predicting greater activity in SMG for auditory but not for audiovisual speech. In addition, connectivity for this pathway only differed across age during audiovisual but not auditory speech, which suggests that this pathway is an important component of a network for processing audiovisual speech. The absence of age differences in BOLD activity in regions previously implicated in audiovisual speech perception is consistent with the interpretation that both children and adults process the visual speech information. However, the network results suggest that less developed networks for integrating auditory and visual speech information in children limit the utility of the visual information for language comprehension.

The neurobiological significance of these network differences might relate to continuing maturational processes of the brain. Brain development continues into early adulthood (Huttenlocher, 1979; Huttenlocher & Dabholkar, 1997; Yakovlev & Lecours, 1967), with regions associated with more primary functions (e.g., sensorimotor cortex, early visual cortex, early auditory cortex) maturing earlier than association cortices. Several studies have suggested that prefrontal association cortices are the last to mature (Huttenlocher & Dabholkar, 1997; Nagel et al., 2006; Sowell, Thompson, Holmes, Jernigan, & Toga, 1999; Sowell et al., 2004), and some research suggests continued maturation of parietal and temporal cortex into adolescence (Giedd et al., 1999; Gogtay et al., 2004). These developmental changes in cortical structure reflect a variety of processes, including a pre-pubertal increase in dendritic and axonal growth and myelination, followed by a post-pubertal process of dendritic pruning and selective cell death (see Paus, 2005 for review). In principle these changes could account for age differences in effective connectivity between inferior

frontal and inferior parietal cortices during audiovisual speech comprehension. In particular, the increased efficiency of the propagation of neural impulses due to myelination might support increased functional integration (Fair, Cohen, et al., 2007; Fair, Dosenbach, et al., 2007). Other factors, however, also contribute to the degree to which regions of the network interact, including coordinated activity ("coactivation") of regions within the network in response to experience (Bi & Poo, 1999; Katz & Shatz, 1996). In fact, both maturational and experiential factors likely contribute to the development of audiovisual speech comprehension, as both processes interact over the course of development (Als et al., 2004; Dawson, Ashman, & Carver, 2000; Edelman & Tononi, 1997; Goldman-Rakic, 1987; Shaw et al., 2008).

The findings we report also raise the question of the potential role for IFGOp/PMv and SMG within a network supporting audiovisual speech comprehension. In the Introduction we suggested that interactions between IFGOp/PMv and posterior temporal/inferior parietal regions relate motor (articulatory) and sensory (auditory and somatosensory) hypotheses about the identity of the speech target (cf. Callan, Callan, Tajima, & Akahane-Yamada, 2006; Callan et al., 2004; Fadiga, Craighero, Buccino, & Rizzolatti, 2002; Pulvermüller et al., 2006; Skipper, van Wassenhove, et al., 2007; Wilson & Iacoboni, 2006). This suggestion is also consistent with the notion that this IFGOp/PMV → SMG pathway is involved in linking articulatory-motor and somatosensory representations during both speech production and perception (Bohland & Guenther, 2006; Callan et al., 2004; Duffau, Gatignol, Denvil, Lopes, & Capelle, 2003; Guenther, 2006; Guenther, Ghosh, & Tourville, 2006; Ojemann, 1992; Skipper, van Wassenhove, et al., 2007). In addition, both of these brain regions show sensitivity to the convergence of auditory and visual speech information. For example, Hasson and colleagues (Hasson, Skipper, Nusbaum, & Small, 2007) found that left IFGOp and SMG were both sensitive to individual differences in the perception of incongruent audiovisual phonemes (i.e., McGurk syllables). That is, although the incongruent phoneme was usually perceived as a fused percept (i.e., auditory /pa/ with video /ka/ is often perceived as /ta/), some people perceived the fused percept more often than others. Further, individuals who failed to perceive the fused audiovisual phoneme, like children, seemed to judge the audiovisual stimulus to be in the phonological category specified by the auditory component, rather than the visual component (cf. Brancazio & Miller, 2005). Finally, Bernstein and colleagues (2008; also see Bernstein et al., 2008) showed that although left STSp, STGp, and precentral gyrus were sensitive to incongruent audiovisual speech sounds, only activation in the left SMG node of this network varied as a function of the incongruity between auditory and visual speech information.

The sensitivity of these regions to audiovisual speech, coupled with their involvement in speech production, supports the hypothesis that the age difference we report reflects a maturation of this functional pathway, in part through experience both producing and perceiving speech. Guenther (Guenther, 2006; Guenther et al., 2006; Tourville, Reilly, & Guenther, 2008) has proposed that during speech production, projections from IFGOp/PMv to posterior superior temporal regions and SMG carry information about the expected auditory and somatosensory traces of produced speech sounds. These expectations are then compared to the actual auditory and somatosensory sensations of the produced sounds, a process that allows rapid articulatory adaptation in response to feedback. Over time children learn to correct speech *production* errors by correcting discrepancies between the motor-speech target (involving IFGOp/PMv) and the actual somatosensory and auditory consequences of the production (involving SMG and posterior superior temporal cortex; Guenther et al., 2006). The mechanism we propose for developing audiovisual speech *perception* is similar; children learn to take advantage of the predictive value of visual

speech information through experience. To the extent that visual speech information is informative for comprehension, this information is incorporated into a mapping system between auditory and somatosensory signals and motor output acquired by perceiving the consequences of the child's own speech productions. Over time, as the mapping between sensory and motor consequences achieves greater consolidation, visual speech information gains predictive value, providing more precise early constraints or "forward models" (Skipper, van Wassenhove, et al., 2007; van Wassenhove et al., 2005) that contribute to the final interpretation of the speech sound.

Importantly, modification of this mechanism may occur for an extended period over the course of development, with changes in both the coupling of one's own *produced* speech and the associated auditory/somatosensory consequences (Plaut & Kello, 1999; Westermann & Miranda, 2004), and changes in the coupling of *visual* speech to its associated auditory consequences (Kuhl & Meltzoff, 1984; Vihman, 2002). Indeed, throughout childhood the speech production system undergoes substantial modification, with anatomical changes and refinement of motor control from infancy to adulthood (Vorperian et al., 2005). These changes have a direct effect on the quality and range of vocal sounds produced by infants and children (Ménard, Schwartz, & Boë, 2004). As the pharynx and vocal tract lengths increase, the mapping of articulatory-to-acoustic and somatosensory representations is updated (Callan, Kent, Guenther, & Vorperian, 2000; Ménard et al., 2004; Ménard, Schwartz, Boë, & Aubin, 2007), in turn impacting the predictive value of visual speech information. In this model, visual speech is not essential, but contributes information when auditory speech targets are ambiguous (for example, when encountering a foreign speaker, or when the auditory signal is degraded by white noise, or by the accompanying MRI scanner noise; Callan et al., 2004; Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007; Ross et al., 2007; Sato, Tremblay, & Gracco, 2009; Schwartz, Berthommier, & Savariaux, 2004). Furthermore, visual information will have greater predictive value in cases where it is most informative (for example, for place of articulation, which might be less clear in the acoustic modality; Binnie et al., 1974), and in cases where it is necessary to supplement degraded auditory speech representations (for example, for children with cochlear implants who show reduced bimodal fusion, and rely more on visual information; Schorr, Fox, van Wassenhove, & Knudsen, 2005).

4.2. Posterior superior temporal sulcus and auditory–visual integration

Left STSp has been hypothesized to be an important cortical region for integration of auditory and visual speech information (Calvert et al., 2000; Campbell, 2008; Okada & Hickok, 2009). Several functional imaging studies have found that STSp shows sensitivity to manipulations of the congruency of auditory and visual speech information (Bernstein et al., 2008; Miller & D'Esposito, 2005; Sekiyama et al., 2003), and responds more strongly to audiovisual than to auditory-only speech (Calvert et al., 2000; Skipper et al., 2005; Wright et al., 2003), to dynamic compared to static visual speech (Calvert & Campbell, 2003), and to visual speech gestures compared to rest (Campbell et al., 2001).

We found that effective connectivity between STSp and a number of regions differed across AV and A language for both children (IFGOp/PMv → STSp) and adults (IFGOp/PMv → STSp; STS → SMG; STSp → STp). For adults, the strength of these connections tended to be greater during audiovisual than auditory-only speech, but the opposite was true for children. To some extent, the findings we report are consistent with the proposed role for STSp in the crossmodal integration of auditory and visual information during speech comprehension (i.e., the analysis of BOLD signal intensity

revealed that STGp/STSp is activated more strongly for audiovisual speech language in children, and for both age groups functional interactions with STSp are modulated by the presence of visual information). However, several findings are not consistent the idea that STSp is a critical region for the integration of audiovisual speech. For example, differences in BOLD signal intensity between the AV and A conditions were not found for adults in STSp. Further, age differences in connectivity were not found for STSp during audiovisual comprehension. Instead, age differences in connectivity were revealed for this region only during the A condition (Fig. 1B; cf. Callan et al., 2003; Hocking & Price, 2008; Jones & Callan, 2003; Ojanen et al., 2005; Olson, Gatenby, & Gore, 2002; Saito et al., 2005 for similar studies reporting no differences between audiovisual and auditory presentation in STSp). Despite some differences in the pattern of the findings, this latter finding is in general agreement with a prior network-level investigation of auditory story comprehension (Karunanayaka et al., 2007). Both studies report that connectivity during auditory story comprehension changes over development. We found that connectivity between posterior inferior frontal/premotor regions and STSp was positive for children, but then became negative for adults. Similarly, Karunanayaka and colleagues found that connectivity between inferior frontal and temporal regions changed with age, although this change was reflected as an increase in connectivity. Notably, the anatomical definition of brain regions in the networks differs between the studies, which might have influenced the connectivity pattern (e.g., the temporal ROI of their network model was functionally defined and included a large part of the anterior superior temporal gyrus and sulcus). However, notwithstanding the lack of correspondence in ROI definition, the difference in the pattern of connectivity changes during auditory story comprehension deserves further exploration. In general, the findings we report do suggest that rather than proposing STSp is a primary binding site for audiovisual speech processing (Campbell, 2008; Okada & Hickok, 2009), the function of STSp must be considered in the context of the network of cortical regions with which it interacts, and the demands of the task. Moreover, electrophysiological evidence suggests integration of auditory and visual information via STSp may occur too late to account for early visual speech effects observed in event-related potential (van Wassenhove et al., 2005) and intracranial event-related potential (Besle et al., 2008) studies. These findings support models emphasizing feed-forward processing of visual speech information, potentially via the application of motor-speech information as we advocate here.

5. Summary

In summary, we used a network modeling approach to examine how the development of audiovisual speech comprehension is reflected by changes in the interactions among brain regions involved in both speech perception and speech production. The analyses we report demonstrated that in children and adults, audiovisual speech comprehension activated a similar fronto-temporo-parietal network of brain regions. Age-related differences in audiovisual speech comprehension were primarily reflected by differences in effective connectivity across the brain regions comprising this network. These findings suggest that the function of a network is not fully characterized by the response of its individual components, but also by the dynamic interactions among them.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bandl.2009.08.005.

References

Als, H., Duffy, F. H., McAnulty, G. B., Rivkin, M. J., Vajapeyam, S., Mulkern, R. V., et al. (2004). Early experience alters brain function and structure. Pediatrics, 113(4), 846–857.

Arbuckle, J. (1989). AMOS: Analysis of moment structures. The American Statistician, 43(1), 66.

Argall, B. D., Saad, Z. S., & Beauchamp, M. S. (2006). Simplified intersubject averaging on the cortical surface using SUMA. Human Brain Mapping, 27(1), 14–27.

Barrett, P. (2007). Structural equation modelling: Adjudging model fit. Personality and Individual Differences, 42, 815–824.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 57(1), 289–300.

Bergeson, T. R., Pisoni, D. B., & Davis, R. A. O. (2005). Development of audiovisual comprehension skills in prelingually deaf children with cochlear implants. Ear and Hearing, 26, 146–164.

Bernstein, L. E., Lu, Z.-L., & Jiang, J. (2008). Quantified acoustic-optical speech signal incongruity identifies cortical sites of audiovisual speech processing. Brain Research, 1242, 172–184.

Besle, J., Fischer, C., Bidet-Caulet, A., Lecaignard, F., Bertrand, O., & Giard, M.-H. (2008). Visual activation and audiovisual interactions in the auditory cortex during speech perception: Intracranial recordings in humans. The Journal of Neuroscience, 28(52), 14301–14310.

Bi, G., & Poo, M. (1999). Distributed synaptic modification in neural networks induced by patterned stimulation. Nature, 401, 792–796.

Binnie, C. A., Montgomery, A. A., & Jackson, P. L. (1974). Auditory and visual contributions to the perception of consonants. Journal of Speech and Hearing Research, 17, 619–630.

Bohland, J. W., & Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. NeuroImage, 32, 821–841.

Brancazio, L., & Miller, J. L. (2005). Use of visual information in speech perception: Evidence for a visual rate effect both with and without a McGurk effect. Perception and Psychophysics, 67(5), 759–769.

Burgund, E. D., Kang, H. C., Kelly, J. E., Buckner, R. L., Snyder, A. Z., Petersen, S. E., et al. (2002). The feasibility of a common stereotactic space for children and adults in fMRI studies of development. NeuroImage, 17, 184–200.

Burnham, D., & Dodd, B. (2004). Auditory–visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. Developmental Psychobiology, 45(4), 204–220.

Byars, A. W., Holland, S. K., Strawsburg, R. H., Bommer, W., Dunn, R. S., Schmithorst, V. J., et al. (2002). Practical aspects of conducting large-scale functional magnetic resonance imaging studies in children. Journal of Child Neurology, 17, 885–889.

Callan, A. M., Callan, D. E., Tajima, K., & Akahane-Yamada, R. (2006). Neural processes involved with perception of non-native durational contrasts. NeuroReport, 17, 1353–1357.

Callan, D. E., Jones, J. A., Callan, A. M., & Akahane-Yamada, R. (2004). Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory–auditory/orosensory internal models. NeuroImage, 22, 1182–1194.

Callan, D. E., Jones, J. A., Munhall, K., Callan, A. M., Kroos, C., & Vatikiotis-Bateson, E. (2003). Neural processes underlying perceptual enhancement by visual speech gestures. NeuroReport, 14, 2213–2218.

Callan, D. E., Kent, R. D., Guenther, F. H., & Vorperian, H. K. (2000). An auditory-feedback based neural network model of speech production that is robust to developmental changes in the size and shape of the articulatory system. Journal of Speech, Language, and Hearing Research, 43, 721–736.

Calvert, G. A., & Campbell, R. (2003). Reading speech from still and moving faces: The neural substrates of visible speech. Journal of Cognitive Neuroscience, 15(1), 57–70.

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Current Biology, 10(11), 649–657.

Campbell, R. (2008). The processing of audio-visual speech: Empirical and neural bases. Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences, 363, 1001–1010.

Campbell, R., MacSweeney, M., Surguladze, S., Calvert, G. A., McGuire, P., Suckling, J., et al. (2001). Cortical substrates for the perception of face actions: An fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning). Cognitive Brain Research, 12(2), 233–243.

Church, J. A., Fair, D. A., Dosenbach, N. U. F., Cohen, A. L., Miezin, F. M., Petersen, S. E., et al. (2009). Control networks in paediatric Tourette syndrome show immature and anomalous patterns of functional connectivity. Brain, 132, 225–238.

Clark, D., & Wagner, A. D. (2003). Assembling and encoding word representations: fMRI subsequent memory effects implicate a role for phonological control. Neuropsychologia, 41, 304–317.

Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. Journal of the American Statistical Association, 83, 596–610.

Cohen, M. S. (1997). Parametric analysis of fMRI data using linear systems methods. NeuroImage, 6(2), 93–103.

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. NeuroImage, 9(2), 179–194.

Daselaar, S. M., Prince, S. E., & Cabeza, R. (2004). When less means more: Deactivations during encoding that predict subsequent memory. NeuroImage, 23, 921–927.

Dawson, G., Ashman, S. B., & Carver, L. J. (2000). The role of early experience in shaping behavioral and brain development and its implications for social policy. Development and Psychopathology, 12(4), 695–712.

Desikan, R. S., Segonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage, 31(3), 968–980.

Desjardins, R. N., Rogers, J., & Werker, J. F. (1997). An exploration of why preschoolers perform differently than do adults in audiovisual speech perception tasks. Journal of Experimental Child Psychology, 66, 85–110.

Desjardins, R. N., & Werker, J. F. (2004). Is the integration of heard and seen speech mandatory for infants. Developmental Psychobiology, 45, 187–203.

Dick, A. S., Goldin-Meadow, S., Hasson, U., Skipper, J. I., & Small, S. L. (2009). Co-speech gestures influence neural activity in brain regions associated with processing semantic information. Human Brain Mapping. doi:10.1002/hbm.20774.

Dodd, B. (1979). Lip-reading in infants: Attention to speech presented in-and-out-of-synchrony. Cognitive Psychology, 11(478–484).

Dodd, B., McIntosh, B., Erdener, D., & Burnham, D. (2008). Perception of the auditory–visual illusion in speech perception by children with phonological disorders. Clinical Linguistics and Phonetics, 22(1), 69–82.

Duffau, H., Gatignol, P., Denvil, D., Lopes, M., & Capelle, L. (2003). The articulatory loop: Study of the subcortical connectivity by electrostimulation. NeuroReport, 14, 2005–2008.

Duvernoy, H. M. (1991). The human brain: Structure, three-dimensional sectional anatomy and MRI. New York: Springer-Verlag.

Edelman, G. M., & Tononi, G. (1997). Commentary: Selection and development: The brain as a complex system. In D. Magnusson (Ed.), The lifespan development of individuals: Behavioral, neurobiological, and psychosocial perspectives: A synthesis (pp. 179–204). Cambridge, UK: Cambridge University Press.

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. European Journal of Neuroscience, 15, 399–402.

Fair, D. A., Cohen, A. L., Dosenbach, N. U. F., Church, J. A., Miezin, F. M., Barch, D. M., et al. (2007). The maturing architecture of the brain's default network. Proceedings of the National Academy of Sciences, 105(10), 4028–4032.

Fair, D. A., Dosenbach, N. U. F., Church, J. A., Cohen, A. L., Brahmbhatt, S., Miezin, F. M., et al. (2007). Development of distinct control networks through segregation and integration. Proceedings of the National Academy of Sciences, 104(33), 13507–13512.

Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis: II: Inflation, flattening, and a surface-based coordinate system. NeuroImage, 9(2), 195–207.

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Segonne, F., Salat, D. H., et al. (2004). Automatically parcelling the human cerebral cortex. Cerebral Cortex, 14(1), 11–22.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proceedings of the National Academy of Sciences, 102(27), 9673–9678.

Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. NeuroImage, 15, 870–878.

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., et al. (1999). Brain development during childhood and adolescence: A longitudinal MRI study. Nature Neuroscience, 2(10), 861–863.

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., et al. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. Proceedings of the National Academy of Sciences, 101(21), 8174–8179.

Goldman-Rakic, P. S. (1987). Development of cortical circuitry and cognitive function. Child Development, 58, 601–622.

Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. Journal of Communication Disorders, 39, 350–365.

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. Brain and Language, 96, 280–301.

Hasson, U., Nusbaum, H. C., & Small, S. L. (2007). Brain networks subserving the extraction of sentence information and its encoding to memory. Cerebral Cortex, 17(12), 2899–2913.

Hasson, U., Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2007). Abstract coding of audiovisual speech: Beyond sensory representation. Neuron, 56, 1116–1126.

Hasson, U., Skipper, J. I., Wilde, M. J., Nusbaum, H. C., & Small, S. L. (2008). Improving the analysis, storage and sharing of neuroimaging data using relational databases and distributed computing. NeuroImage, 39(2), 693–706.

Hayasaka, S., & Nichols, T. E. (2003). Validating cluster size inference: Random field and permutation methods. NeuroImage, 20(4), 2343–2356.

Hocking, J., & Price, C. J. (2008). The role of the posterior superior temporal sulcus in audiovisual processing. Cerebral Cortex, 18, 2439–2449.

Hockley, N. S., & Polka, L. (1994). A developmental study of audiovisual speech perception using the McGurk paradigm. The Journal of the Acoustical Society of America, 96, 3309.

Huttenlocher, P. R. (1979). Synaptic density in human frontal cortex: Developmental changes and effects of aging. Brain Research, 163, 195–205.

Huttenlocher, P. R., & Dabholkar, A. S. (1997). Regional differences in synaptogenesis in human cerebral cortex. The Journal of Comparative Neurology, 387, 167–178.

Jaccard, J. (1998). Interaction effects in factorial analysis of variance. Thousand Oaks, CA: Sage.

Johnstone, T., Ores Walsh, K. S., Greischar, L. L., Alexander, A. L., Fox, A. S., Davidson, R. J., et al. (2006). Motion correction and the use of motion covariates in multiple-subject fMRI analysis. Human Brain Mapping, 27, 779–788.

Jones, J. A., & Callan, D. E. (2003). Brain activity during audiovisual speech perception: An fMRI study. NeuroReport, 14, 1129–1133.

Kang, H. C., Burgund, E. D., Lugar, H. M., Petersen, S. E., & Schlaggar, B. L. (2003). Comparison of functional activation foci in children and adults using a common stereotactic space. NeuroImage, 19(1), 16–28.

Karunanayaka, P. R., Holland, S. K., Schmithorst, V. J., Solodkin, A., Chen, E. E., Szaflarski, J. P., et al. (2007). Age-related connectivity changes in fMRI data from children listening to stories. NeuroImage, 34, 349–360.

Katz, L. C., & Shatz, C. J. (1996). Synaptic activity and the construction of cortical circuits. Science, 274(5290), 1133–1138.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. Science, 218, 1138–1140.

Kuhl, P. K., & Meltzoff, A. N. (1984). The intermodal representation of speech in infants. Infant Behavior and Development, 7, 361–381.

Lachs, L., Pisoni, D. B., & Kirk, K. I. (2001). Use of audiovisual information in speech perception by prelingually deaf children with cochlear implants: A first report. Ear and Hearing, 22(3), 236–251.

MacLeod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. British Journal of Audiology, 21, 131–141.

Massaro, D. W. (1984). Children's perception of visual and auditory speech. Child Development, 55, 1777–1788.

Massaro, D. W., Thompson, L. A., Barron, B., & Laren, E. (1986). Developmental changes in visual and auditory contributions to speech perception. Journal of Experimental Child Psychology, 41, 93–113.

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

McIntosh, A. R., & Gonzalez-Lima, F. (1992). Structural modeling of functional visual pathways mapped with 2-deoxyglucose: Effects of patterned light and foot shock. Brain Research, 578, 75–86.

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

McIntosh, A. R., Grady, C. L., Ungerleider, L. G., Haxby, J. V., Rapoport, S. I., & Horwitz, B. (1994). Network analysis of cortical visual pathways mapped with PET. The Journal of Neuroscience, 14(2), 655–666.

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. Current Biology, 17, 1692–1696.

Ménard, L., Schwartz, J.-L., & Boë, L.-J. (2004). Role of vocal tract morphology in speech development: Perceptual targets and sensorimotor maps for synthesize French vowels from birth to adulthood. Journal of Speech, Language, and Hearing Research, 47, 1059–1080.

Ménard, L., Schwartz, J.-L., Boë, L.-J., & Aubin, J. (2007). Articulatory–acoustic relationships during vocal tract growth for French vowels: Analysis of real data and simulations with an articulatory model. Journal of Phonetics, 35, 1–19.

Miller, L. M., & D'Esposito, M. (2005). Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. Journal of Neuroscience, 25(25), 5884–5893.

Nagel, B. J., Medina, K. L., Yoshi, J., Schweinsburg, A. D., Moadab, I., & Tapert, S. F. (2006). Age-related changes in prefrontal white matter volume across adolescence. Developmental Neuroscience, 17, 1427–1431.

Nichols, T., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. Human Brain Mapping, 15(1), 1–25.

Noll, D. C., Cohen, J. D., Meyer, C. H., & Schneider, W. (1995). Spiral K-space MR imaging of cortical activation. Journal of Magnetic Resonance Imaging, 5(1), 49–56.

Ojanen, V., Mottonen, R., Pekkola, J., Jaaskelainen, I. P., Joensuu, R., Autti, T., et al. (2005). Processing audiovisual speech in Broca's area. NeuroImage, 25, 333–338.

Ojemann, G. (1992). Localization of language in frontal cortex. In P. Chauvel, A. V. Delgado-Escueta, E. Halzgren, & J. Bancaud (Eds.), Advances in neurology (pp. 361–368). New York: Raven Press Ltd.

Okada, K., & Hickok, G. (2009). Two cortical mechanisms support the integration of visual and auditory speech: A hypothesis and preliminary data. Neuroscience Letters. doi:10.1016/j.neulet.2009.01.060.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia, 9(1), 97–113.

Olson, I. R., Gatenby, J. C., & Gore, J. C. (2002). A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. Cognitive Brain Research, 14(1), 129–138.

Parrish, T. B., Gitelman, D. R., LaBar, K. S., & Mesulam, M.-M. (2000). Impact of signal-to-noise on functional MRI. Magnetic Resonance in Medicine, 44, 925–932.

Patterson, M. L., & Werker, J. F. (2003). Two-month old infants match phonetic information in lips and voice. Developmental Science, 6(2), 191–196.

Paus, T. (2005). Mapping brain maturation and cognitive development during adolescence. Trends in Cognitive Sciences, 9(2), 60–68.

Pekkola, J., Ojanen, V., Autti, T., Jääskelainen, I. P., Mottonen, R., & Sams, M. (2006). Attention to visual speech gestures enhances hemodynamic activity in the left planum temporale. Human Brain Mapping, 27(6), 471–477.

Petersson, K. M., Reis, A., Askelöf, S., Castro-Caldas, A., & Ingvar, M. (2000). Language processing modulated by literacy: A network analysis of verbal repetition in literate and illiterate subjects. Journal of Cognitive Neuroscience, 12(3), 364–382.

Pienaar, R., Fischl, B., Caviness, V., Makris, N., & Grant, P. E. (2008). A methodology for analyzing curvature in the developing brain from preterm to adult. International Journal of Imaging Systems and Technology, 18(1), 42–68.

Plaut, D. C., & Kello, C. T. (1999). The emergence of phonology from the interplay of speech comprehension and production: A distributed connectionist approach. In B. MacWhinney (Ed.), The emergence of language (pp. 381–415). Mahwah, NJ: Erlbaum.

Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. Proceedings of the National Academy of Sciences, 103(20), 7865–7870.

Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. Perception and Psychophysics, 59(3), 347–357.

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. Cerebral Cortex, 17(5), 1147–1153.

Saito, D. N., Yoshimura, K., Kochiyama, T., Okada, T., Honda, M., & Sadato, N. (2005). Cross-modal binding and activated attentional networks during audio-visual speech integration: A functional MRI study. Cerebral Cortex, 15, 1750–1760.

Sato, M., Tremblay, P., & Gracco, V. L. (2009). A mediating role of the premotor cortex in phoneme segmentation. Brain & Language, 111(1), 1–7.

Schmahmann, J. D., & Pandya, D. N. (2006). Fiber pathways of the brain. Oxford, England: Oxford University Press.

Schmithorst, V. J., Holland, S. K., & Plante, E. (2006). Cognitive modules utilized for narrative comprehension in children: A functional magnetic resonance imaging study. NeuroImage, 29(1), 254–266.

Schorr, E. A., Fox, N. A., van Wassenhove, V., & Knudsen, E. I. (2005). Auditory–visual fusion in speech perception in children with cochlear implants. Proceedings of the National Academy of Sciences, 102(51), 18748–18750.

Schwartz, J.-L., Berthommier, F., & Savariaux, C. (2004). Seeing to hear better: Evidence for early audio-visual interactions in speech identification. Cognition, 93, B69–B78.

Sekiyama, K., & Burnham, D. (2008). Impact of language on development of auditory–visual speech perception. Developmental Science, 11(2), 306–320.

Sekiyama, K., Kanno, I., Miura, S., & Sugita, Y. (2003). Auditory–visual speech perception examined by fMRI and PET. Neuroscience Research, 47(3), 277–287.

Sekiyama, K., & Tohkura, Y. (1993). Inter-language differences in the influence of visual cues in speech perception. Journal of Phonetics, 21, 427–444.

Shaw, P., Kabani, N. J., Lerch, J. P., Eckstrand, K., Lenroot, R., Gogtay, N., et al. (2008). Neurodevelopmental trajectories of the human cerebral cortex. The Journal of Neuroscience, 28(14), 3586–3594.

Shulman, G. L., Fiez, J. A., Corbetta, M., Buckner, R. L., Miezin, F. M., Raichle, M. E., et al. (1997). Common blood flow changes across visual tasks: II. Decreases in cerebral cortex. Journal of Cognitive Neuroscience, 9, 648–663.

Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., & Small, S. L. (2007). Speech-associated gestures, Broca's area, and the human mirror system. Brain and Language, 101(3), 260–277.

Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., & Small, S. L. (2009). Gestures orchestrate brain networks for language understanding. Current Biology, 19, 1–7.

Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2005). Listening to talking faces: Motor cortical activation during speech perception. NeuroImage, 25(1), 76–89.

Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2006). Lending a helping hand to hearing: Another motor theory of speech perception. In M. A. Arbib (Ed.), Action to language via the mirror neuron system (pp. 250–285). Cambridge University Press.

Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. Cerebral Cortex, 17, 2387–2399.

Small, S. L., & Nusbaum, H. C. (2004). On the neurobiological investigation of language understanding in context. Brain and Language, 89(2), 300–311.

Solodkin, A., Hlustik, P., Chen, E. E., & Small, S. L. (2004). Fine modulation in network activation during motor execution and motor imagery. Cerebral Cortex, 14, 1246–1255.

Sowell, E. R., Thompson, P. M., Holmes, C., Jernigan, T. L., & Toga, A. W. (1999). In vivo evidence for post-adolescent brain maturation in frontal and striatal regions. Nature Neuroscience, 2(10), 859–861.

Sowell, E. R., Thompson, P. M., Leonard, C. M., Welcome, S. E., Kan, E., & Toga, A. W. (2004). Longitudinal mapping of cortical thickness and brain growth in normal children. The Journal of Neuroscience, 24, 8223–8231.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. Journal of the Acoustical Society of America, 26, 212–215.

Summerfield, Q. (1979). Use of visual information in phonetic perception. *Phonetica, 36*, 314–331.

Tamnes, C. K., Østby, Y., Fjell, A. M., Westlye, L. T., Due-Tønnessen, P., & Walhovd, K. B. (2009). Brain maturation in adolescence and young adulthood: Regional age-related changes in cortical thickness and white matter volume and microstructure. *Cerebral Cortex*. doi:10.1093/cercor/bhp118.

Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition, 108*, 850–855.

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage, 39*(1429–1443).

van Linden, S., & Vroomen, J. (2008). Audiovisual speech recalibration in children. *Journal of Child Language, 35*, 809–822.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences, 102*(4), 1181–1186.

Vihman, M. M. (2002). The role of mirror neurons in the ontogeny of speech. In *Mirror neurons and the evolution of brain and language* (pp. 305–314). Amsterdam: John Benjamins.

Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., & Yandell, B. S. (2005). Development of vocal tract length during early childhood: A magnetic resonance imaging study. *Journal of the Acoustical Society of America, 117*, 338–350.

Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. F. (2007). Visual language discrimination in infancy. *Science, 316*(5828), 1159.

Werker, J. F., Frost, P. E., & McGurk, H. (1992). La langue et les lèvres: Cross-language influences on bimodal speech perception. *Canadian Journal of Psychology, 46*(4), 551–568.

Westermann, G., & Miranda, E. R. (2004). A new model of sensorimotor coupling in the development of speech. *Brain and Language, 89*, 393–400.

Wightman, F., Kistler, D., & Brungart, D. (2006). Informational masking of speech in children: Auditory–visual integration. *Journal of the Acoustical Society of America, 119*, 3940–3949.

Wilson, S. M., & Iacoboni, M. (2006). Neural responses to non-native phonemes varying in producibility: Evidence for the sensorimotor nature of speech perception. *NeuroImage, 33*(1), 316–325.

Wilson, S. M., Molnar-Szakacs, I., & Iacoboni, M. (2008). Beyond superior temporal cortex: Intersubject correlations in narrative speech comprehension. *Cerebral Cortex, 18*, 230–242.

Wright, T. M., Pelphrey, K. A., Allison, T., McKeown, M. J., & McCarthy, G. (2003). Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cerebral Cortex, 13*(10), 1034–1043.

Yakovlev, P. I., & Lecours, A.-R. (1967). The myelogenetic cycles of regional maturation of the brain. In A. Minkowsky (Ed.), *Regional development of the brain in early life* (pp. 3–70). Oxford, UK: Blackwell.

Yarkoni, T., Speer, N. K., Balota, D. A., McAvoy, M. P., & Zacks, J. M. (2008). Pictures of a thousand words: Investigating the neural mechanisms of reading with extremely rapid event-related fMRI. *NeuroImage, 42*, 973–987.